ARIZONA READING FIRST

ANNUAL EVALUATION REPORT

Year 2, 2003 - 2004

Submitted to
Arizona Department of Education

Conducted by

Arizona Prevention Resource Center
Arizona State University
542 E. Monroe, Bldg. D
Phoenix Arizona 85004

and

Northwest Regional Educational Laboratory
101 SW Main, Suite 500
Portland, Oregon 97204

October, 2004

ACKNOWLEDGEMENTS

An evaluation that covers as complex and far-reaching a project as Arizona Reading First cannot be accomplished without the collaborative efforts of many people. We would like to acknowledge the many individuals who answered our endless questions, located data, and tolerated intrusion into their already busy schedules. Chief among those are the staff at the Arizona Department of Education (ADE) itself, led by Marie Mancuso, who were indispensable sources of information themselves and furthermore modeled a support for evaluation that made collection of data from other sources easier.

Comprehensive, intensive site visits require considerable effort, time, and cooperation from a range of people. We would like to acknowledge the site visitors who worked so hard to ensure the collection of high-quality data: Christopher Emge, Maria Moratto, Angela Aguayo and Lynn Katz from the APRC as well as Kari Nelsestuen, Carolyn Moilenen, and Theresa Deussen from NWREL. Furthermore, the quality of the observations was boosted immensely by the willingness of administration and staff at two schools to let us come and pilot test observation protocols in their classrooms. Many thanks to the principals, coaches, and teachers at Gilbert Heights Elementary in Portland, Oregon, and at Whitman Elementary in Mesa, Arizona, for their gracious accommodation of the teams of evaluators who descended upon them.

We especially need to single out the principals, coaches, and teachers at the Arizona Reading First schools that were selected for site visits this year and give them our thanks. On top of the hard work they did for the implementation of the grant, they also demonstrated great hospitality to their site visitors, despite the additional demands of time and energy involved. We hope they will feel that the time they spent has yielded findings that will ultimately be of use to everyone involved in the project.

The County Reading Specialists moved well outside their zone of comfort when they took on the task of not simply cheerleading and encouraging growth at their schools but also using a protocol to attempt to measure progress in implementation. We are grateful for their willingness to take on a challenging new role as well as their time and insights share in a focus group conducted by Wendy Wolfersteig last spring.

Contributions to data analysis and writing of findings deserve recognition as well. Elizabeth Autio at NWREL provided analyses of survey findings from the 2003 Summer Academy; the 2004 follow-up findings were analyzed by Steve Capobres with assistance from Richard Nagasawa, both of APRC. Wendy Wolfersteig, along with Christopher Emge, Maria Moratto and Richard Nagasawa at APRC synthesized the findings from the County Reading Specialist focus group. Kari Nelsestuen at NWREL handled both the analysis and the writing of the findings from the implementation checklist. Theresa Deussen coordinated the analyses of massive amounts of site visit data by a team of very thoughtful evaluators at NWREL, including Kari Nelsestuen, Elizabeth Autio, Angela Roccograndi, Chari Smith, John Towner, Elizabeth Vale and Judy Van Scoter. Kari Nelsestuen and Elizabeth Autio also played very large roles in triangulating qualitative and survey data for this final report. Assistance with the DIBELS analysis was provided by NWREL to Wendy Wolfersteig and Maria Moratto who produced the tables and narrative. Wendy Wolfersteig, Christopher Emge, and Richard Nagasawa at APRC analyzed and reported on the AIMS and Stanford 9 assessment data.

We also want to acknowledge the contributions of staff members who provided the some of the technical, administrative, and other support functions that made the analyses possible. Forrest Stevens of NWREL created and maintained the database for the site visit data. At APRC, all the checklist and survey data were scanned by Yu-Yu (Candy) Choi; Anji Pon helped with phone interviews and Shelley Kuklish assisted with data tables and analysis. Both the NWREL and APRC teams spent countless hours developing and editing the reports – thanks to all.

TABLE OF CONTENTS

1. Executive Summary 1

2 Introduction 11

3. Methodology 17

4. Student Assessments I: DIBELS by Grade Level 36

5. Student Assessments II: A Closer Look at DIBELS School Level Variation 77

6. Student Assessments III: AIMS and Stanford 9 Reading Results 98

7. Professional Development 113

8. Leadership and School-Level Structures 141

9. Data and Assessment Systems 166

10. Instruction 172

References 209

Appendix A: DIBELS Demographics by Measure 210

Appendix B: DIBELS NWF as Predictor for ORF, First Grade 231

Appendix C: DIBELS First Grade NWF and ORF by School 237

Appendix D: AIMS Scores by School 2002, 2003, and 2004 302

Appendix E: Stanford 9 Scores by School 2003 and 2004 311

Appendix F: Training Satisfaction Instruments (DIBELS and LETRS trainings)

Appendix G: Summer Academy Pre and Post-Surveys

Appendix H: Spring Follow-up Surveys

Appendix I: Implementation Checklists, February and May

Appendix J: County Reading Specialist Survey and Focus Group with Protocol

Appendix K: Questionnaire with Basic Information about Their Reading Program

Appendix L: Interview Protocols

Appendix M: Observation Protocol and Rating Rubric

Appendix N: District Interview Protocol

Appendix O: State Interview Protocol

CHAPTER 1
EXECUTIVE SUMMARY

Overall Findings

The Arizona Department of Education was awarded a federal Reading First grant in order to improve primary reading instruction and ensure that all students can read at grade level by the end of third grade. The past academic year (2003-04) marked the first year of implementation in the 63 Arizona schools that were awarded Reading First subgrants. This executive summary presents a brief overview of the findings of the external evaluation of that first year.

Arizona Reading First accomplished a great deal in its first year of implementation in the schools, including the following:

- delivery of focused professional development to about 1,200 teachers and specialists;
- provision of high-quality, targeted technical assistance in order to strengthen instruction and build school leadership;
- high levels of buy-in /commitment from County Reading specialists, principals, and coaches;
- adoption of scientifically-based core reading programs, for the most part with a high commitment to fidelity; and
- establishment of school-wide systems for the administration of assessments and entry of results into a data system that provided immediate, user-friendly feedback.

At the same time, many schools met with significant challenges as they implemented the project in their schools. Some of those that appeared at many - certainly not all - schools included:

- ensuring a high level of understanding and use of assessment data to guide instructional decisions;
- classroom management and instructional strategies that did not always maximize student engagement and the effective use of time;
- little development of school-wide systems to deliver targeted interventions to students who needed additional instruction in reading; and
- uneven teacher buy-in to the program, especially among third-grade teachers.

At the end of the first year, student assessment results indicated that there was a substantial increase in the percentage of kindergarten students at benchmark on early reading skills (+44%). At the first grade level, results showed a small increase (+7%) in the percentage of students at benchmark. There were very small increases in second and third grade as well, although when compared to declines in the percentage at benchmark at comparison schools, the small increases look positive.

On the AIMS (third grade) and Stanford 9 (second and third grade) assessments, the performance of Reading First students did not change, compared to the previous year and continued to be lower than the state average. This may be due both to the short amount of time that the new reading programs were in place and the general lack of comprehensive interventions for those students who were furthest behind. Of course, there was variation among schools and some Reading First saw larger improvements than average in their first year.

Report Highlights

In this first year of school-level intervention, schools faced very demanding expectations. They adopted new core reading programs which they were expected to utilize without incorporating components of whatever programs they were used to teaching from in the past. They learned a new reading assessment (the DIBELS) and were to make decisions about grouping, instruction, and the delivery of interventions based on assessment results. Collaboration around reading school-wide was supposed to increase, with the creation of a Reading Leadership Team and an improvement plan, regular grade-level meetings, and near daily walk-throughs from the principal, who was supposed to become an instructional leader for the school. While this was happening, the principal and a full-time reading coach were participating in intensive and on-going professional development which they were to carry back to their schools to support the implementation of Reading First. Not surprisingly, some schools were able to meet these expectations better than others, and some expectations were better met by more schools than others.

The external evaluation analyzed student assessment scores to determine whether there were changes in reading performance over the course of the first year. In addition, using a variety of evaluation methods and tools, the evaluation looked at the degree to which schools were able to implement what Reading First expected of them and what else they needed from state project staff in order to meet the challenges that arose.

Highlights from the different sections of the evaluation report follow below. All of these points, along with the way data were collected and analyzed, are explained in much greater depth in the body of the report.

Student Performance on the DIBELS Assessment

- All of the Reading First grade level groups showed *positive* change.

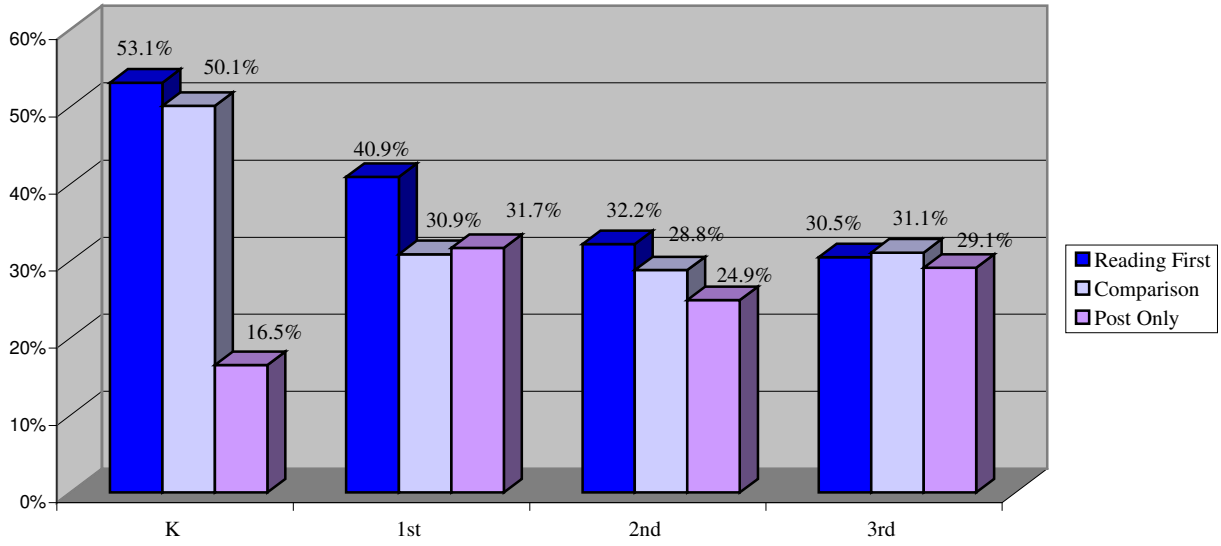
- The percent of students at "benchmark" *increased* for each of the grade levels over the percent at benchmark at the beginning of the year.

- Kindergarten difference in the benchmark category from the beginning to end of the year was the most dramatic (44 percent).

- While first graders saw a small increase (7.4 percent) in those at benchmark from the beginning to end of year, the second and third graders differences in percent at benchmark (3.5 percent and 2.7 percent respectively) were almost flat.

- The percent in the benchmark category at the end of the year was highest in kindergarten at over 50 percent, with the first graders at 41 percent and only about one-third of second and third graders at benchmark at the end of the year.

Percentage of Students at Benchmark on the DIBELS, Spring 2004



Reading First Performance and Comparison Schools

There were two groups of schools against which Reading First schools were compared: the "Comparison" group (schools that already administered the DIBELS assessment but did not have Reading First programs) and the "Post-only" group (schools that did not use the DIBELS assessment, in which state project staff administered post-tests in the spring of 2004).

- The percentage of kindergarten students at benchmark improved more among Reading First than among comparison schools. Both the Reading First and the Comparison schools had substantially more kindergarten students at benchmark at the end of the year than did schools in the post-only group.

- In the first, second and third grades, the percentage of Reading First students at benchmark increased a little from beginning to end of the year. In contrast, at Comparison schools the percentage at benchmark declined from beginning to end. By the end, Reading First schools had more students at benchmark than either the Comparison or the Post-only schools.

- Overall performance on the DIBELS assessment, as measured by the percentage of students at benchmark, varied considerably by school. It is crucial that the review of project-wide trends noted be supplemented by an awareness of this school-level variation.

Student Performance on the AIMS

- On the AIMS reading test, the third graders in Arizona had an overall weighted mean score that was lower in 2004 than in 2003.

- The mean scores of third graders in 2004 in the Reading First (505) and Comparison schools (507) continued to be lower than the mean score of third graders in the state (519).

- From 2003 to 2004, the percentage of the Arizona third graders who "Meet" or "Exceed" the standard *decreased* by 5 percent. Reading First and comparison schools also saw declines, of 5.8 percent and 11.5 percent, respectively.

- Between 2003 and 2004, the percentage of student scoring at the "Falls Far Below" proficiency level, increased by 6.9 percent at Reading First schools, 11.6 percent at comparison schools, and 4.0 at all Arizona schools across the state.

Student Performance on the Stanford 9 Assessment

- On the Stanford 9 reading assessment, the mean percentile score of the Reading First second graders increased by 2 points in 2004 over 2003, whereas the mean score of second graders in the comparison schools declined by 2 points.

- The mean percentile scores for third grade students at the Reading First and comparison school third graders decreased slightly (by less than one point) from 2003 to 2004; both were below the state average.

- For both second and third graders in all Arizona schools, the mean percentile scores on the Stanford 9 reading test did not change from 2003 to 2004 and were 11 to 15 points higher than the scores for the Reading First schools for both years.

- The second- and third-grade Reading First percentile scores were two to six points higher than those of the comparison schools in both 2003 and 2004.

Professional Development

Reading First professional development encompassed both the training provided by the state to teachers (the Summer Reading Academies) and to principals and coaches in the form of monthly meetings, as well as district- and building-level training provided to teachers.

- Principals spoke positively about the professional development provided to them by the state. About half of them said that these trainings had caused them to gain a better general understanding of the Reading First program and a specific understanding of the five essential reading components.

- In interviews, some principals suggested additional training in motivating staff, "working with resistant teachers" and "building buy-in." Others suggested further training in how to use data to drive instruction, specifically "how that would look on a day to day basis."

- Coaches overwhelmingly indicated that the training provided to them by the ADE had been very helpful and useful. Conversely, a small number of coaches and specialists complained that some of the monthly meetings were redundant or not applicable to their position. About half of coaches said that they would like additional training in coaching methods.

- Almost all assessment coordinators had positive feedback regarding the professional development that they had received through the state; they found it "useful," "excellent" and

were "amazed with the quality." They requested more training in various aspects of working with data especially help with data presentations and communicating data to teachers.

- There was little evidence indicating substantial training in the use of assessment data or strategies for English Language Learner students although these were challenging issues for the schools.

- Teaching staff indicated that, of all professional development that they received, they were most impressed by the professional development provided at the school-level by the reading coach.

- Vocabulary, comprehension, and fluency were the top three components pointed out as areas in which a need was indicated for additional training for teachers. Some also expressed an interest in receiving additional training in classroom management, student engagement, using data and assessment to guide instruction, and working with English language learners.

- In interviews, principals and coaches overwhelmingly praised the ADE for the support and technical assistance it provided them in the first year of the grant.

- County Reading Specialists provided a wide range of services based on the schools' needs and were largely viewed as very supportive, although a few schools had less positive experiences. They played an important role in supporting reading assessment by assisting with DIBELS training, administration, and data interpretation.

Leadership and School-Level Structures

Some of the biggest changes prompted by Reading First were the creation and use of new institutions within schools and new ways for teachers, coaches, and principals to work together to improve reading instruction for their students.

The Role of LEAs

- The most common role for LEAs to take during this first year of implementation was grant administration and management. Some LEAs also provided technical assistance and material support. Those that provided technical assistance reported that this was more time-consuming than anticipated.

- Representatives from most LEAs indicated that the Reading First project had an impact on many or all the elementary schools in their district, not just the Reading First schools. Some LEAs had provided reading training for non-Reading First schools, or encouraged schools to use the DIBELS assessment, while others had made scientifically-based reading materials more broadly available.

- Most of the schools visited expressed appreciation for the support of the grant from their LEAs, support that came in the form of meetings, sharing of knowledge, or arrangement of substitutes to facilitate training attendance.

- About a third of schools described some problem or frustrations in their work with LEAs on Reading First, sometimes because of poor communication or a lack of understanding about what the LEA role should look like.

Leadership and Key Roles

- Staff at most schools described their principals as "very involved" in the implementation of the Reading First grant at their school. Most principals met regularly with teachers and the coach, and participated in reviewing and sharing assessment data. Almost all principals conducted classroom observations and attended grade-level meetings, although these were the lowest-rated items on the implementation checklist.

- Coaches were most likely to list observing classrooms, mentoring teachers, conducting demonstration lessons or other professional development, and assessing students or managing assessment data as the tasks which took most of their time. Most coaches were also fulfilling the other roles thus nearly half reported a shortage of time to complete all the tasks as a big enough concern to mention in the interviews.

- Coaches reported that they were confident in their role as coach and that this confidence increased as the year progressed. A substantial number faced resistance from teachers ranging from building trust with teachers unaccustomed to classroom observations to struggling with openly hostile teachers who did not buy into the RF program.

- Most schools had assessment coordinators, although these were often part-time positions. Responsibility for a range of other tasks besides coordinating Reading First assessments, including training teachers about assessments, working closely with the Reading Leadership Team, and managing progress monitoring, meant that assessment coordinators listed time constraints as their number one challenge.

- Most assessment coordinators were excited about the growing interest in and reliance upon assessment data at their schools, although a few reported that low assessment scores made them personally unpopular at their schools.

- A few schools had highly functional Reading Leadership Teams that actively reviewed data together, made decisions and worked to ensure program fidelity. Many schools had teams that served primarily as information sharing structures; their role was very much still evolving. On the other end of the spectrum, a few schools had teams that met rarely or not at all.

Communication, Collaboration, and Support for Reading First

- According to surveys, support for Reading First was fairly high at the start of the grant; however, support declined slightly over the year. Data revealed a larger decline in support among third-grade teachers compared to other grades.

- These levels of support for Reading First were, for the most part, evenly spread across schools. A handful of schools reported very positive experiences with high teacher buy-in;

conversely, a handful of schools had relatively poor experiences, with teachers displaying "serious" resistance.

- Across the board, schools reported improvements in communication and collaboration over the past year; however, the degree of improvement varied greatly from school to school.

- In general, those schools with higher levels of buy-in for Reading First at the beginning of the year indicated greater successes in increasing communication and collaboration, compared to schools that had lower levels of initial buy-in.

- One of the most commonly cited challenges across the evaluation was teacher resistance; developing ways to draw teachers into the program – and identifying and working with the schools that struggle most in this area – may be an area in which support from the ADE coming year could be beneficial.

Data and Assessment Systems

- Schools reported that they used the DIBELS assessment, as well as core program assessments to "inform instruction," by which they primarily meant screening, progress monitoring, and assigning students to groups.

- Advantages of the assessment systems were that data served as a basis for teacher collaboration and planning and communication with parents about their child's progress.

- Most schools appreciated and valued the DIBELS assessment system, particularly for the timeliness of the information they received. At a few schools, there were concerns about the applicability of DIBELS to new ELL students, to older students who read at very low levels, or about contradictory results from different assessments.

Instruction

Ultimately, all of the other changes promoted under Reading First were designed to alter student experience within the classroom in order to improve student reading ability.

Core Reading Programs

- Overall satisfaction with the different core reading programs at the visited schools was high, regardless of which program was adopted. Teachers and specialists rated their core programs high on the five essential components such as providing explicit instruction in comprehension strategies and exposing students to vocabulary in a variety of contexts.

- Although overall satisfaction was high, most interviewees pointed out gaps and flaws in their core programs (core program manuals were not user-friendly for teachers; writing was not integrated in the program; materials were too difficult for ELL students; or certain components such as comprehension were stronger in some grades than in others). Also, some schools complained about difficulties in obtaining all the materials the publishers had promised them.

- Schools knew that "fidelity to the core program" was important to Reading First staff. Many schools tended to interpret this very strictly, as a charge to teach only and exactly what was in the book. Some schools articulated a balance between the scope and sequence of their core program and thoughtful adaptation to student needs, while a few schools appeared to adopt a looser interpretation of fidelity ("the spirit, not the letter, of the program"). The core program was used in almost all observations conducted.

Classroom Instruction

- Site visitors observed instruction in all of the five essential components of reading. Kindergarten lessons tended to focus on phonemic awareness and phonics. In first grade, these were also the primary areas of instruction, although work in fluency and comprehension were observed as well. In second and especially in third grade, comprehension was the most commonly observed focus of instruction, followed by vocabulary and fluency.

- Phonemic awareness and phonics lessons tended, for the most part, to be fast-paced and to promote student engagement. In the area of vocabulary instruction, site visitors witnessed both highly engaging, active lessons, and slower, less interesting reviews of vocabulary words. In the area of comprehension, it was more common to see teachers asking students to answer simple recall questions than to think more deeply about the meaning of text. These observations mesh with statements from coaches, who often reported that vocabulary and comprehension instruction were the areas they felt needed most improvement in their schools.

- Because observations were comparatively short, one-time visits, it cannot be assumed that failure to observe a particular strategy meant that teachers never used such a strategy. What observations do suggest, when combined into a big picture, is the degree to which certain practices have become customary in some classrooms. For example,
    - In about half of observed lessons, teachers regularly monitored student understanding and adjusted the pace of instruction to fit student needs.
    - In about the same proportion of observations, site visitors observed that the teacher provided clear and appropriate feedback to students as they worked.
    - Explicit modeling was noted in about half of all observations.
    - In getting the most use possible out of available instructional time, observers noted some examples of excellent classroom management that moved students efficiently from one task to another, but in many cases slow transitions or failure to have materials ready ahead of time meant there was wasted time in classrooms.
    - There were many instances in which teachers did not make use of think-pair-share, partner reading, or other strategies that would have meant more students were practicing reading skills at the same time.

Interventions

- During the first year of implementation, state project staff placed primary emphasis on implementation of the core program and many schools reported leaving the establishment of an intervention system for the next year.

- Some schools had built an intervention program. Often they reported using materials attached to their core reading program, with varying degrees of success. Only a few schools had well-developed systems, complete with appropriate supplementary materials and regular progress-monitoring.

- Many schools struggled with the scheduling and staffing demands of putting together enough interventions for all the students who needed it. Schools that were able to draw on volunteers or paid outside tutors or had teachers work extra hours to provide interventions were very grateful for the additional staffing.

- School staff requested training and technical assistance to increase buy-in for creating a school-wide intervention plan, to better target interventions to specific needs, to better interpret/use DIBELS data, and to determine which materials best serve which purposes.

Meeting the Needs of English Language Learner Students

For most Reading First schools, the needs of English Language Learners were real and pressing concerns.

- At the end of the first year of implementation, about a third of teachers indicated a need for some assistance in learning how to adapt instruction to meet the specific needs of students who were not native speakers of English.

- Likewise, principals, coaches, and districts all expressed a need for more information and more support on working effectively with ELL students.

- Opinions about how well the core reading programs served the needs of ELL students varied tremendously.

Recommendations

More detailed descriptions of the following recommendations are provided in the body of the report. Briefly, however, the evaluation findings summarized above suggest that the following areas deserve attention in the coming year:

- High priority should be given to the establishment and effective functioning of comprehensive, school-wide intervention systems in order to provide support to struggling students.

- Those schools which have struggled most with implementation need more frequent and targeted support from state project staff (this is already underway).

- Particular emphasis should be placed on the enhancement of instruction in vocabulary and comprehension; continued work to improve classroom management and maximize student engagement is needed.

In addition, based on participant feedback and evaluation findings, there are certain topic areas that should be covered in the state's provision of on-going professional development:

- At every level, school staff need to deepen their understanding of what it means to use data to drive instruction.

- Coaches and County Reading Specialists have requested additional assistance in learning how to perform their demanding roles effectively and in how to build teacher buy-in.

- Principals need continued support to perform effective and meaningful classroom observations and walk-throughs regularly.

- At some schools, staff need support in understanding how to work collaboratively on planning and/or use of data.

CHAPTER 2
INTRODUCTION

Reading First – The Federal Level

Reading First is a federal initiative authorized by the amendments to Title I, Part B, Subpart 1 of the Elementary and Secondary Education Act through the *No Child Left Behind Act* of 2001. The Act authorizes the U.S. Department of Education to award Reading First grants to states, which in turn award subgrants to eligible districts and schools to utilize the findings of scientifically-based reading research to improve K-3 reading instruction and student learning.

This act built on earlier work begun in 1998, under the *Reading Excellence Act* (REA), which was an amendment to Title II of the Elementary and Secondary Education Act (ESEA) of 1965. Reading First differs from REA both in the amount of funding to states, which was significantly increased, and in the higher level of guidance and stricter requirements and accountability accompanying the grants.

The ultimate purpose of Reading First is to ensure that all children read at grade level by the end of third grade. In support of this goal, funds are provided to states to support comprehensive programs to improve reading instruction at selected Reading First schools as well as more broadly in the state.

Under guidance from the U.S. Department of Education, statewide Reading First programs are to promote scientifically-based reading instruction focused on the five essential components of reading: phonemic awareness, phonics, reading fluency, reading comprehension strategies, and vocabulary development. Instruction in these essential components should be systematic and explicit, that is, it should follow a logical and coherent scope and sequence and explicit modeling and explanation. Furthermore, instruction should be targeted to students' individual needs, with the provision of additional instruction and support to struggling students. The regular use of valid and reliable reading assessments provides information on those students needs. Teachers should receive high quality professional development in order to understand and effectively implement changes in the core reading program, instructional strategies, and use of assessment data. Efforts should be coordinated within and across grades, requiring higher levels of collaboration and strong leadership at the school level. Thus the federal Reading First office, refers to the four "pillars" supporting Reading First: professional development, leadership, assessment, and instructional programs and materials.

Reading First – The State Level

The state of Arizona was awarded a Reading First grant in 2003 and invited subgrant applications from eligible schools districts (Arizona refers to a districts as a Local Education Associations – LEA). The LEAs were selected on the basis of the following criteria: (1) student reading achievement scores on the AIMS pass rate for 3rd grade;
(2) pass rate of 75 percent or less; and (3) whether or not the LEA exhibited one of the following items--poverty rate, number of schools identified for school improvement, and location of the LEA in an empowerment zone. The screening resulted in 138 LEAs that were eligible for to apply for Reading First.

At the time of application, state project staff informed districts that the grants would be accompanied by both support and specific expectations, including the following:

- Selection and implementation of core reading program materials from a list of suggested research-based materials or demonstration that a selected program met the criteria for scientifically-based reading research.
- Selection and implementation of research-based reading interventions for students who need them.
- Attendance of all K-3 staff at the state's Summer Reading Academy each year.
- Hiring of a full-time reading coach to provide mentoring, coaching, training and demonstration lessons.
- In many instances, hiring of a full- or part-time assessment coordinator to implement a coherent assessment system and promote the thoughtful use of assessment data.
- Creation of a Reading Leadership Team (RLT) to guide the design of a K-3 reading delivery system.
- Attendance of reading coaches, principals, and district coordinators at monthly trainings.
- Use of approved valid and reliable assessments in fall, winter, and spring for K-3, analyses of results and use of data to make reading improvement decisions.
- Site visits and use of feedback from independent evaluators, as well as state and federal Reading First administrators.

The award of Reading First subgrants to 63 schools in 26 districts was announced in May 2003. Table 1, below, lists these schools by district and includes the total enrollment in grades K-3, the percent of students who received FRL, and the percent of students who were English Language Learners (ELL).

Table 2-1
Arizona Reading First School Demographics (2003-2004)

	K-3 Enrollment	Percent Free Reduced Lunch %	Percent ELL
School, by District			
Alhambra			
Westwood Primary	1125	82	*
Sevilla	843	94	*
Andalucia	1060	98	62
Casa Grande			
Ironwood	341	69	20
Mesquite Elementary	403	71	19
Coolidge			
West School	698	78	16
Crane			
Pueblo	426	64	28
Rancho Viejo	641	96	63
H.L. Suverkrup	378	81	41
Valley Horizon	532	62	33
Glendale			
American School	533	73	25
William C. Jack	1094	72	*
Isaac			
Mitchell Elem.	397	90	73
P.T. Coe	614	91	85
J.B. Sutton Elem.	414	96	71
Liberty			
Rainbow Valley	204	46	5
Maricopa			
Phoenix Pappas	331	98	15
Mesa			
Hawthorne	471	60	14
Holmes	496	82	33
Lowell	525	92	67
Roosevelt	377	65	15
Whitman	435	69	15
Whittier	359	61	17
Nogales			
Bracker Elementary	153	83	46
Challenger	387	82	62
Mitchell Elementary	401	85	77
Page			
Desert View	331	87	52
Parker			
LePera School	148	89	31
Pendergast			
Westwind	706	62	58
Pendergast	416	66	34

	K-3 Enrollment	Percent Free Reduced Lunch %	Percent ELL
Red Mesa			
Red Mesa	181	78	83
Round Rock	84	95	85
Roosevelt			
J.R. Davis School	276	94	75
Jorgensen School	325	88	58
M.O. Bush School	374	73	31
Southwest School	369	69	29
Sunland School	332	87	45
T.G. Barr School	339	81	50
Safford			
Lafe Nelson School	268	62	11
Somerton			
Desert Sonora	277	86	79
Orange Grove	252	87	74
Tierra Del Sol	511	88	78
Stanfield			
Stanfield Elementary	352	95	50
Sunnyside			
Craycroft Elementary	395	90	40
Drexel Elementary	441	95	89
Summit View	366	90	61
Tempe			
Curry	398	56	*
Evans School	318	63	22
Laird School	352	78	51
Tolleson			
P.H. Gonzales	433	79	38
Tucson			
C.E. Rose	351	95	*
Davidson	281	89	30
Lynn/Urquides	539	92	42
Menlo Park	231	97	47
Pueblo Gardens	202	96	33
Roberts	347	96	53
Washington			
Mountain View	533	93	70
Shaw Butte	463	74	46
Wickenburg			
MacLebbab	562	42	14
Willcox			
Willcox	277	67	36
Yuma			
Gwyneth Ham School	390	88	38
Palmcroft School	455	79	24
Roosevelt	437	84	62
AVERAGE	428	79.5	

* For a few schools, the percentage of English Language Learners was not available.

Both large and small schools received Arizona Reading First grants. The largest, in terms of K-3 enrollment, were the three schools in the Alhambra district and William C. Jack Elementary in Glendale; these schools served between 800 and 1,150 students. Three other schools had over 600 K-3 students. On the other end of the spectrum, four schools had fewer than 200 students. Three of these were rural, and one was a charter school.
The vast majority of participating schools had an enrollment ranging from 200 to 600 students in grades K-3, with the average enrollment at 428 students.

In all but two schools, a majority of students were eligible to receive Free/Reduced-Price Lunch; the average was about 80 percent. Every participating school reported that they received Title 1 funds.

The percentage of ELL students ranged from zero to 85 percent of the student body, with 85 percent of the schools serving an ELL population of at least 20 percent. The district with the highest percentage of ELL students was Red Mesa's two schools (85%) that are located on the Navaho reservation. Other schools with a high ELL population included Drexel (81%), Desert Senora (79%) and Tierra Del Sol (78%).

The Evaluation

The Arizona Department of Education Reading First project contracted with the Arizona Prevention Resource Center (APRC) to conduct the statewide evaluation, beginning with the 2002-2003 year. The first year of the project was spent in planning, preparation, and early training. In May 2003, APRC contracted with the Northwest Regional Educational Laboratory to assist with the evaluation. This is the second annual report, but the first one to focus on the school-level implementation of Reading First.

The desired outcome of Reading First is to see an improvement student reading assessments. Specifically, the DIBELS test scores directly reflect improvement in students' reading in grades K-3; it measures the concepts taught by the reading curriculum. There should also be an increase in AIMS and Stanford 9 achievement test scores; however, large improvements were not expected in the first years as these tests measure only second and third grade students who may not have begun their reading instruction with a science-based reading curriculum.
Of short-term importance were changes in the "four pillars" as these aimed to impact the educational system at the teacher, school, district and state levels. These efforts in professional development, leadership, assessment systems, and instruction in science-based reading were implemented as the schools began the 2003-2004 year. The goal was to measure the implementation, process, throughputs and system improvements on an on-going basis in order to make data-based decisions and changes in implementation as needed.

Six broad questions define the focus of the evaluation. These evaluation questions were addressed using a range of approaches and instruments which are described in Chapter 3 (Methodology):

1. Professional Development: How effective was the professional development approach in helping teachers and administrators acquire knowledge and skills about phonemic awareness, phonics, fluency, vocabulary development, and reading comprehension, and transfer the knowledge and skills to their classroom instruction?
2. Knowledge Transfer to Students: How effective was Arizona's Reading First initiative in increasing students' knowledge and abilities related to phonemic awareness, phonics, fluency, vocabulary development and reading comprehension?
3. Transfer of Knowledge in the Classroom: To what extent are teachers incorporating reading assessments into their classrooms and using the results of the assessments to change their instructional approaches and address students' learning needs?
4. Knowledge Transfer to Students: How well are K-3 Arizona students meeting the standards for performance in reading as measured by the Arizona Instrument to Measure Success (AIMS), the state assessment, and to what extent is performance improving over time?
5. Capacity Building - Support System: How effective is the system of support for schools and districts to help all key stakeholders to contribute to the improvement of students' reading performance and sustain improved performance over time?
6. Capacity Building – Leadership: To what extent has Reading First help develop instructional leadership in coaches, principals, and LEAS?

While the evaluation addressed components of all of these questions, for the first year of school implementation, some questions had greater relevance than others. Emphasis was placed on learning what went smoothly in early implementation and what aspects were more challenging. In addition, comprehensive analyses of the DIBELS student assessment were undertaken. Trends in student performance can best be examined in the future, when the schools have implemented Reading First for multiple years.

Organization of This Report

This report examines the status of project implementation and early student assessment outcomes at the end of Year 2, which is actually the Year 1 of implementation in the schools. It is organized to along the following logic. First, it presents student assessment data (DIBELS, AIMS and Stanford 9), in order to summarize the status of schools at the end of their first year, as well as report on the growth they made in the first year.

Next, the report moves on to look more closely at the implementation of Reading First – information which provides the context and, perhaps, the reasons for the student assessment results presented first. Implementation findings are organized according to the "four pillars" of Reading First: professional development, leadership, assessment systems, and instruction. Each of these pillars references an evaluation question.

In order to facilitate use of this lengthy report, each chapter begins with highlights from that chapter, many of which are also in the Executive Summary. Relevant conclusions and/or recommendations are contained in each section of the report as appropriate.

CHAPTER 3
METHODOLOGY

Overview

A multi-method strategy was used to evaluate the Reading First project outcomes and processes. This chapter discusses the many data collection instruments. Chapters 4 through 10 discuss the findings in an integrated manner.

In order to address both the evaluation questions and to document achievements and challenges in Reading First, evaluators utilized a number of different methodologies and instruments to collect a great deal of information. Whenever possible, evaluators collected information from more than one source (such as from principals as well as teachers) and/or from more than one instrument (interviews as well as surveys), in order to "triangulate," that is, to look at things from more than one point of view.

The DIBELS, AIMS and Stanford 9 reading scores were used to assess the desired outcome of improvement in student reading achievement. Results from more than one instrument were used to assess the areas of the "four pillars" of Reading First: professional development, leadership, assessment systems, and instruction. These areas address the implementation, process, throughputs and system improvements

The following instruments, discussed in this chapter (see Appendixes F to O), were used during this first year evaluation:

- Training satisfaction instruments (DIBELS and LETRS trainings): addressed how satisfied participants were with various aspects of their learning at these sessions.
- Surveys (Summer Academy pre and post-test and spring follow-up): these instruments contained question on attitudes, instructional practices, experiences, and the knowledge questions related to learning about the five areas of reading.
- Implementation Checklist (February and May): Completed by the County Reading Specialists, these 76+ item instruments assessed progress of implementation of RF in the districts, schools and classrooms.
- County Reading Specialist focus group and survey: questions designed to quantify and qualify the experiences of the CRS in their work during this first year.
- Site visits – observations and interviews: These observations, interviews, and visits followed protocols in order to make judgments across site as to what was occurring at the school and classroom level with administrators, teachers and students.
- District and state interviews: District phone interviews and a personal interview with the Director of Reading First sought additional information on the context, degree, substance and acceptance of Reading First.

The remainder of this chapter describes each of these instruments in detail, as well as the response rates obtained and any limitations or cautions about the data collected via one of the instruments.

DIBELS Assessment

The DIBELS test is a valid and reliable indicator of early literacy development and predictive of later reading proficiency. It is able to identify students who are not progressing as expected. The result of DIBELS testing can also be used to evaluate individual student development and provide grade-level feedback on instructional objectives. The DIBELS measures were devised for K-3 students on the basis of the skills that are fundamental to later reading success. DIBELS assess the 5 essential components of reading instruction: 1) phonological awareness (ISF & PSF); (2) phonics (NWF); (3) vocabulary (WUF); (4) Fluency (ORF), and (5) comprehension (Retell). DIBELS was developed at the Institute for the Development of Educational Achievement at University of Oregon.

For the purpose of this evaluation, DIBELS test scores were used to answer a specific research question:

- Knowledge Transfer to Students
  How effective was Arizona's Reading First Initiative in increasing students' knowledge and abilities related to phonemic awareness, phonics, fluency, vocabulary development, and reading comprehension?

Matching Procedure
The DIBELS evaluation called for a pretest-posttest comparison group design to measure the impact of the RF program on the DIBELS reading performance test. In the Reading First schools, all K-3 pupils were exposed to science-based reading instruction in the classroom. The analysis examined the change in DIBELS scores between the pre and posttests (beginning of year and end of year).

The evaluation design was a pre- and posttest single group design used to measure the changes over time in the perceptions, attitudes, and behavior of teachers, coaches, specialist, and principals regarding the RF program and its implementation in the schools.
This meant having two comparison groups: a Comparison group in which students took both the pre and post tests, and a Post-test only comparison group to eliminate the potential effects of the use of DIBELS in the comparison schools.

The RF schools were matched with all comparison schools in order to make the schools comparable on key variables. To create the match, the principals of the Reading First schools were first asked to identify the schools in the LEA that best represented a "match" for their schools. The rationale here was that the principals were likely to have first-hand knowledge of the demographic characteristics of the local area. The principals' list thus served as a guide in the matching process. The Reading First schools were then matched with eligible non-Reading First schools on the following key variables: Title I, percent poverty, and percent of students who fell below the AIMS third Grade reading standard. Finally, the schools within the matched LEAs were further matched on the basis of: Percent Free/Reduced lunch; Stanford 9 second and third grade reading levels; and ethnicity.

The principals of the matched comparison schools were invited to volunteer their schools to serve as comparison schools to the RF schools. At issue here was the degree to which the

schools elected to participate. ADE contacted the schools asking for participation, and most agreed. However, when calls were placed to schools in the spring, some schools "forgot" that they had agreed to participate, others had changed personnel, and some were already administering DIBELS. Thus, the numbers of matched schools were significantly reduced so other schools, which were not as closely matched, were selected to replace the original matched schools. This resulted in sixteen matched schools (Table 3.1) that agreed to release their data to ADE and the evaluation team for comparison purposes.

Table 3.1
Comparison Schools' Core Reading List

RF LEA	LEA	Comparison Schools – Pre-test & Post-test	Core Reading Program	SBRR Core
YES	1. Alhambra	Granada Primary School	Scholastic Literacy Place	NO
	2. Murphy	Alfred F. Garcia School	Open Court	YES
YES	3. Roosevelt	Ed & Verna Pastor Elementary	Universal Learning--Voyager	YES
YES	4. Roosevelt	Martin Luther King Jr. Elementary	McGraw-Hill Reading	YES
	5. Bullhead City	Coyote Canyon School	McGraw-Hill Reading	YES
	6. Holbrook	Park Elem (K-2) and Hulet Elem (3)	Houghton Mifflin	YES
	7. Pendergast	Desert Horizon	No Core	NO
	8. Hyder	Dateland Elementary	"Sing, Read, Write and Spell" (K-2), MacMillian McGraw Hill" (3rd)	NO (K-2) YES (3rd)
		Comparison Schools - Post test Only		
	1. Buckeye	Buckeye Elementary	Richard C. Owens' The Literacy	NO
	2. Mohave Valley	Mohave Valley Elementary	Scott Foresman	NO
YES	3. Mesa	Longfellow	Harcourt Trophies	YES
	4. Picacho	Picacho School	Scott Foresman.	NO
	5. Phoenix	Paul Laurence Dunbar	MacMillian-McGraw Hill	YES
	6. Toltec	Toltec Elementary	No reading Program	NO
	7. Ft. Thomas Unified	Ft Thomas Elementary	SRA Reading Mastery	NO
YES	8. Yuma	O.C. Johnson	Harcourt Trophies	YES

The Comparison group that administered both the pre-post test consisted of four schools within RF LEAs and four from nearby LEAs. Of these eight, five schools had instituted a recommended science-based reading curriculum (SBRR) with another school using it for third grade only. The second comparison group of eight schools, the Post only group, were schools in similar LEAs (two were in RF LEAs). The key was that these students had no prior exposure to DIBELS before taking the test at the end of the year. Three of these schools were using a SBRR curriculum.

DIBELS test administration
The DIBELS test was administered to K-3 students in the Reading First schools to obtain pretest (baseline) measures prior to the initiation of the SBRR programs in the fall of 2003. Indeed, for the Reading First schools, DIBELS was administered three times during the school year to students in grades K, 1, 2, and 3 (several schools choose to administer the test four times). In the beginning (fall) of 2003, the pretest was also administered to K-3 students in eight-matched comparison schools (some of these schools also administered the test in the middle of the year). In the end (spring) of 2004, the DIBELS test was administered to students in the RF and all 16 (both groups) of comparison schools.

School staff administered the DIBELS test to students in the RF schools and the eight Comparison (pre-post) schools. The ADE staff administered the DIBELS exam to a sample of students in the second group of eight comparison schools, the Post only group, during April and May of 2004. The sample of students, approximately 30 per grade level per school, was selected by alternating a combination of criteria including school name, grade level, classroom, teachers' first or last name, and beginning or end of alphabet.

Table 3. 2
DIBELS Tests by Group

Groups	Pretest	Posttest
Reading First	Yes	Yes
Comparison Group (pre and post test)	Yes	Yes
Post Only Comparison	No	Yes

These scores were captured in the DIBLES database maintained by the University of Oregon. Schools could run out their own reports of individual student and school or district data. Further, because schools districts submitted a permission form to the University of Oregon for release of their data, both ADE and the evaluation team had access to these databases and directly downloaded the data.

Analysis
The data were downloaded from the DIBLES website by grade level for each of the three groups: RF, Comparison (prepost) group and Post only comparison group. Further, data for the RF and the Comparison group were downloaded to ***match*** students who had taken both the pre and post tests at the beginning of the year (fall) and again at the end of the year (spring). However, scores on each individual test were not captured for each student. Although only students who were

matched (had both pre and post test scores) were used in this sample, the number of students per individual test varied from test to test, and from pre to post test. The middle of year test administration was not considered

Table 3.3
Number of Students per Group per Grade Level*

Number of Students	Kindergarten	1st grade	2nd grade	3rd grade
Reading First	5,580	5,749	5,620	5,328
Comparison	880	939	873	682
Post Only	253	189	213	196

*although only students who were matched (had both pre and post test scores) were used in this sample, the number of students per individual test varied

Once downloaded, the data were then recoded into new variables to reflect the "at risk," "some risk," or "low risk" categories, and then were further recoded into new variables to show the score combinations that reflect the Instructional Support Recommendation of "intensive," strategic" or "benchmark."

Descriptive statistics were reported for all the test measures. This included means, median, percentiles, standard deviation for each measure. For the categories, frequencies and percentages were calculated. Change scores were also computed for the test measures that had both pre and posttest scores, and for the percent who changed within the benchmark category at the end of the year.

Analysis of Covariance
The statistical method of choice in analyzing the data generated by the pretest and posttest control group design is analysis of covariance, in which the posttest measures are compared using pretest means as the covariate. For first, second and third grade data, one-way analysis of covariance was used.

The independent variable, group, included two levels: Reading First group and the prepost Comparison group. The dependent variable was the DIBELS reading score.
For purposes here, it was assumed that the homogeneity of slopes relating the covariate to the dependent variable did not differ significantly.

The quarterly report showed the analysis for the pretest and posttest comparison data using one-way analysis of variance of the pre-post "change scores." The analysis of variance (ANOVA) was used to analyze the pretest and posttest *change scores*, as shown in (a), (b), and (c) above (i.e., the differences between the differences). The residue (d - d' = D) thus gives the measure of the impact of the program in question. This strategy of analysis was presented in the quarterly report for reasons noted; however, since the statistical method of choice is analysis of covariance, those results were presented here in Chapter 4 as additional supportive data to the earlier analysis.

The logic of the comparisons in the pre-and posttest control group design is consistent with an analysis of the pre and posttest ***change scores***. The design and analysis are shown in Figure 3.1.

Figure 3.1
Analysis of DIBELS Reading Change Scores

Design			**Analysis**
R	O_1	X O_2	(a) $O_2 - O_1 = d$
	O_3	O_4	(b) $O_4 - O_3 = d'$
	where O = score		(c) $d - d' = D$ or effect of Intervention

X = intervention, R = randomization

Limitations
When comparing across groups, a word of caution is advised. These groups were matched based upon school and demographic characteristics. However, there is the possibility of some "spillover" effect of RF, especially within those schools in RF districts. RF LEAs have been encouraged to share professional development and other activities to improve reading. Also, the RF emphasis on the 90 minute block reading time may have occurred in other schools regardless of the reading program as the ADE has encouraged additional emphasis on reading through AZ Reads. This might be likely especially at the kindergarten level with the heightened awareness in the state of the impact of kindergarten for academic as well as social learning along with the movement toward all day kindergarten in many schools.

There are also other factors crucial to the examination of differences, which have to do with the size of the sample and population. In this study, both the RF schools and the Comparison school groups comprise the entire population (not a sample) whereas the Post only group is a sample of the students. Further, with a number over 5,000 for the RF populations, there is a greater probability that findings will be significant.

Arizona's Instrument to Measure Standards (AIMS)

For the purpose of this evaluation, AIMS and Stanford 9 test scores were used to answer a specific research question:

- Knowledge Transfer to Students
  How well are K-3 Arizona students meeting the standards for performance in reading as measured by the Arizona Instrument to Measure Success (AIMS), the state assessment (Stanford 9), and to what extent is performance improving over time?

The AIMS test is a standards-based test which provides information regarding the progress of Arizona's students toward mastering Arizona's reading, writing, and mathematics standards. It is a criterion-referenced assessment. The AIMS test shows how well students are mastering leaning goals and how they compare with other children statewide. The test consists of a combination of multiple-choice, short answers, and essay items. For reading, the student is expected to apply reading strategies, evaluate literary elements, evaluate technical manuals, and analyze classic and contemporary literature selections. The AIMS test is not timed. The standards for the third grade AIMS Reading test scale scores are given below:

Table 3.4
AIMS 2004 Reading Performance Level Scale Scores

	Scale Score	Raw Score	%
Falls Far Behind (FFB)	300-473	0-21	
Approaches	474-499	22-29	51%
Meets Standard	500-546	30-38	70%
Exceed Standard	547-700	39-43	91%

For the purpose of this report, the third grade AIMS reading data were downloaded from the ADE website. The Reading First schools and the Comparison group of sixteen schools were sorted from this data-download of all schools. When scores are shown for all Arizona schools, the RF and Comparison groups were included to reflect the same scores shown by the ADE for all Arizona schools. The data were analyzed to look at changes across years in the four categories: "falls far behind," approaches," "meets standard," and "exceeds standard." Also, for each of the groups and across the groups, the differences were calculated in the mean scores percent change as shown in Figure 3.2

Figure 3.2
Analysis of AIMS Reading Scores

		Spring 2003		Spring 2004	% Change
	Schools				
[1]	Reading First	O_1	X	O_2	d_1
[2]	Comparison Group	O_3		O_4	d_2
	Difference:	[4]		[5]	[3]

Stanford Achievement Test (SAT-9)

The Stanford Achievement Test (SAT 9 or Stanford 9) is a timed, norm-referenced, multiple-choice test in reading, math and language. For these purposes, the focus is on the Stanford 9 second and third grade reading scores percentile ranks. The reading test assess comprehension of three types of reading material: textural, recreational, and functional.

The Stanford 9 data are reported as percentile ranks for the school level compared with others in the same grade based on the 1995 norm testing group. A percentile rank reflects the typical student's performance at the school compared to the norming group for that grade and subject area. Thus, if the school score is 39, it means that the average student at this school scored better than 39% of the students in the 1995 norming group. Schools with ranks reported near the 50th percentile indicate that the typical student performance on the test is about average when compared with other students of the same grade level. Higher percentile ranks reflect better performance.

For the purpose of this report, the second and third grade SAT 9 reading data were downloaded from the ADE website. The Reading First schools and the Comparison group of sixteen schools were sorted from this data-download of all schools. When scores are shown for all Arizona schools, the RF and Comparison groups were included to reflect the same scores shown by the ADE for all Arizona schools. The data were analyzed to examine changes across years and across the groups; the differences were calculated in the mean scores percent change as shown in Figure 3.3

Figure 3.3
Analysis of Stanford 9 Reading Percentile Ranks

	Schools	Spring 2003		Spring 2004	% Change
[1]	Reading First	O_1	X	O_2	d_1
[2]	Non-Reading First	O_3		O_4	d_2
	Difference:	[4]		[5]	[3]

Training Satisfaction Instruments (DIBELS and LETRS trainings)

For the purpose of this evaluation, DIBELS and LETRS training satisfaction instruments were used to answer a specific research question:

- Professional Development and Knowledge Transfer to Teachers
  How effective was the professional development approach in helping teachers acquire knowledge and skills about phonemic awareness, phonics, fluency, vocabulary development, and reading comprehension, and transfer the knowledge and skills to their classroom instruction?

Even more specifically, these professional development activities address the issue of knowledge transfer to trainers and support staff, who then in turn, support the teachers and student learning.

The *Dynamic Indicators of Basic Early Reading Skills* (DIBELS) two training workshops were held in August and September 2003 by ADE and helped to guide district assessment teams in the use and interpretation of the DIBELS assessment. The first workshop focused on administration of the DIBELS. Workshop participants then returned to their schools to administer the assessment. The second workshop, held two weeks later, focused on interpretation of DIBELS data and participants were invited to bring their schools' results that had been gathered during the administration between the two workshop dates, to this training to learn how to interpret these data.

Second, the *Language Essentials for Teachers of Reading and Spelling* (LETRS) Institutes fall cycle 2003 consisted of the nine modules in three books: Book One, Foundations of Reading Instruction, Modules 1, 2, and 3 in September; Book Two, Vocabulary, Fluency and Comprehension, Modules 4, 5, and 6 in November; and Book Three, Teaching and Assessing Beginning Reading and Spelling, Modules 7, 8, and 9 in December. This fall cycle was specifically for Reading First participants.

The feedback surveys were conducted to measure general impressions of the training sessions. Topics on the DIBELS surveys included use of the DIBELS assessment, preparation to administer the DIEBLS, and satisfaction with the training. The LETRS feedback surveys focused on content and format of the training itself.

Feedback surveys were collected from 473 participants in the DIBELS Administration session and 228 participants from the DIBELS Interpretation session. Participants were primarily reading coaches, assessment specialists, principals, and other school and district administrators.

Participants were asked to rate their understanding of the DIBELS on a scale of zero to six (with a zero denoting no prior understanding and a six denoting a thorough understanding) at three points: (1) prior to all Reading First training, (2) prior to the DIBELS Administration training, and (3) after the DIBELS Administration training.

The participants of the LETRS Institutes were asked to rate, on a scale of zero (low/ "poor") to six (high/ "superior") four items that related to the ***format*** of the three Institutes. Two items were rated that related to the content, i.e., the extent to which the contents enhanced participants'

ability to use SBRR and were relevant to their work (on a scale of zero (low/"not at all" to high/"to a great extent").

Surveys: Attitudes, Practices, Experiences and Knowledge Questions (Summer Academy pre and post-test and spring follow-up surveys)

For the purpose of this evaluation, Surveys (pre and post) from the Summer Reading Academy and spring follow-up were used to answer a specific research question:

- Professional Development
  How effective was the professional development approach in helping teachers acquire knowledge and skills about phonemic awareness, phonics, fluency, vocabulary development, and reading comprehension, and transfer the knowledge and skills to their classroom instruction?
  How effective was the professional development approach in transferring knowledge to Teachers? How effective was the professional development approach in transferring knowledge to trainers and support staff?

Pre-surveys were administered to participants on their first morning at the Academy. The initial survey included items on attitudes, instructional practices, collaboration, use of assessment data, and coaching. Teachers, reading coaches, reading specialists, and principals received similar surveys with slight differences appropriate to their role in the school. Some questions may have been worded slightly differently on one survey than on another; for ease of reporting on a single table, the wording was chosen to best reflect the meaning of each item.

In addition, knowledge assessments were added to the surveys. Only the knowledge questions were repeated as a post-test measure on the last day of the Academy to ascertain the level of learning regarding the increase in knowledge of the essential reading components. The pre and post-test single group design was used to evaluate the effects of the academy on teachers' *knowledge* of reading instruction in order to determine the effectiveness of the reading academy in transferring knowledge to teachers. In this design, every participant received the potential benefits of the program. It called for a measure of change over the duration of the program. If scores differed between the pre and post-tests, then the difference can be attributed to the program (i.e., the program is said to have an effect on the outcome). The same knowledge questions were posed to the group in the spring follow-up survey to assess the retention of the knowledge.

The pre-surveys were distributed in the morning General Session. The post-test surveys were distributed in the breakout sessions, prior to the last general session of the Academy. Unfortunately, the original protocol designed for this purpose was not followed. Hence, a more formal protocol to distribute the pre and post-test surveys was established in the breakout sessions under the control of the session leaders or presenters. For this purpose, the session leaders were presented with written explanations to frequently asked questions (FAQ). The leaders were asked to read the explanations in the breakout sessions so that the participants would have a better understanding of the purpose of the surveys and thereby cooperate fully and

honestly answer the survey questions. Hence, one FAQ handout was written for the pre-test; another was written for the post-test.

At the first Academy, pre-surveys were administered to participants as they sat in one large group for the welcome to the Academy. Evaluators noted that there was a great deal of talking and sharing of information during the survey, causing concern about the validity of the surveys, especially the knowledge assessments. For this reason, at all subsequent Academies, participants were broken into their smaller grade-level groupings prior to receiving the surveys. They were also provided with more complete information about the purpose of the survey and the evaluation more broadly. This format was repeated for the remaining Academies.

Across the six Summer Academies, pre and post survey data were collected from 1,391 participants including 987 teachers, 206 specialists, 93 coaches, and 105 principals.

The follow-up surveys were administered in several ways. The principals completed their surveys at a principals' meeting on April 6; principals from 24 schools did not complete surveys. The coaches completed their surveys at a coaches training meeting on Wednesday or Thursday, April 14 or 15; coaches from 11 schools did not complete the survey. The coaches were also given packets of surveys to distribute at their schools to the specialists and teachers. Returned via postage paid mail, surveys were not received for specialists from 28 schools and from teachers at 7 schools.

There was some confusion over who was classified as a specialist and this may have led to a lower return rate. Further, some respondents completed a different survey on the follow-up, as when IDs were matched, the type of survey completed (teacher, specialist, coach) did not always match the original type. In the spring, 1,046 follow-up surveys were received and it appeared as if there were 842 teachers, 75 specialists, 89 coaches and 40 principals. Only 618 of these cases matched IDs from the pre-test data (many surveys had blank ID numbers); this group was used to report the matched pre to follow-up attitude results without specifying by group type in order to avoid any mis-identification by group.

Table 3.5
Surveys: Pre, Post and Follow-up

Survey Name	When Administered	Items
Pre-test Surveys	Before Summer 2003 Reading Academy	Attitudes, Practices, Experiences and Knowledge Questions
Post-test Surveys	After Summer 2003 Reading Academy (four days later)	Attitudes, Practices, and Experiences
Follow-up Surveys	In Spring 2004 (approx. nine months later)	Attitudes, Practices, Experiences and Knowledge Questions

For the attitude, instruction and experience items, the pretest and follow-up responses were computed into percentages. The attitude, instruction and experience analysis in the Quarter 4, 2004 report focused on the changes between *all* the pre-test responses and *all* follow-up responses that were received and were not matched by individual or school. However, in this

Annual Report, the respondents were matched yielding 618 teachers, coaches, and specialists with matched scores for whom comparative analyses are presented.

For the knowledge pre-post test, initially the pre- to post-test scores were compared with 959 respondents; at the end of the year, the post-test measure was compared to the follow-up measure for 603 matched respondents.

Implementation Checklist

The Implementation Checklist is a unique instrument, designed to address whether or not various aspects of the Reading First program were implemented in the schools and classrooms. The Checklist questions cover the four pillars (professional development, leadership, assessment systems, and instruction) and provide information on four of the research questions:

- Professional Development and Knowledge Transfer to Teachers: How effective was the professional development approach in helping teachers acquire knowledge and skills about phonemic awareness, phonics, fluency, vocabulary development, and reading comprehension, and transfer the knowledge and skills to their classroom instruction?
- Transfer of Knowledge in the Classroom: To what extent are teachers incorporating reading assessments into their classrooms and using the results of the assessments to change their instructional approaches and address students' learning needs?
- Capacity Building - Support System: How effective is the system of support for schools and districts to help all key stakeholders to contribute to the improvement of students' reading performance and sustain improved performance over time?
- Capacity Building – Leadership: To what extent has Reading First help develop instructional leadership in coaches, principals, and LEAS?

The starting point for what eventually became the Arizona Reading First implementation checklist was an instrument designed by Jo Robinson and used by grantee schools in the Washington Reading Excellence Act project in 2001-03. The document was first revised to better address the specifics of Reading First in Arizona, which meant dropping some items, adding others, and rearranging items into nine broad categories:

- Leadership
- Assessment
- Intervention strategies
- Instruction
- Communication and collaboration
- Professional development
- Reading coach
- Environment
- District support

Seven of the items in the checklist were yes/no items while the remaining items were asked on a four-point scale where "1" indicated "no implementation," "2" indicated "scant implementation," "3" indicated "some implementation," and "4" indicated "fully implemented." Many of the

items were descriptive in nature, asking if and how much activities were occurring in the school (e.g., if Reading Leadership Teams were meeting quarterly, if the core reading program was being used by all teachers). Some items, especially those related to instruction, addressed issues of the quality of implementation.

A series of discussions between the ADE and the evaluation team determined that it would be the County Reading Specialists (CRSs) who would use the checklist, and that the information would be helpful both to the ADE for monitoring purposes and to the evaluation team to have a view of implementation to validate their own observations from site visits and interviews. Initial changes in the wording and in the rating scale were made, first by the ADE itself, and second in a collaborative process involving a trainer from WestEd, a member of the evaluation team, and the CRSs. In a daylong meeting in October 2003, the CRSs, the trainer and evaluator met again to clarify some of the items. During this meeting, the meaning of the questions and how/what the ratings would mean were discussed and agreed upon, and although everyone took notes, no formal written explanations were provided at that time. It was also recognized that the information would be shared among the pertinent parties for process improvement.

After the first checklist was administered in January and February 2004, evaluators summarized findings in a quarterly report to the ADE which was shared with CRSs at a monthly meeting. After discussion of the instrument, several additional revisions were made including the addition of written descriptors that were discussed and agreed upon in October 2003. The descriptors were added to the instrument in hopes of further standardizing its use by CRSs, an issue of validity raised after the first checklist was administered.

In both February and May, Implementation Checklists representing all 63 RF schools were completed.

Limitations

The ADE and evaluators hoped that the addition of item descriptions before the second administration would improve standardization across CRS' ratings. However, because the added descriptions changed the instrument, comparisons over time (from February to May) must be interpreted with caution. The refined definitions may have made it more difficult - or easier - for a school to receive a high score in May.

Change-over-time results in the Implementation Checklist results must also be interpreted with caution because of the change in the audience and use of the instrument. Before the second administration, the ADE clarified that all checklists would be shared with school principals. In some schools, principals became much more involved in the rating process, providing input before ratings were finalized. In at least a few cases, CRSs noted that they gave higher ratings (e.g., a "3" instead of a "2") than they might have given if their ratings had been confidential.

Another issue of concern is reliability – in this case, whether different CRSs used the rating scale in a consistent fashion. The added item descriptors before the second administration was one way to address this concern, but there were still many factors that could have contributed to different ratings among the CRSs.

By themselves, the Checklist ratings should be considered as not perfect measures; caution should be used and especially in comparing one school to another. Importantly, as cited in this report, the Implementation Checklist data results were compared with survey and qualitative results for a deeper understanding of program implementation in Reading First schools.

County Reading Specialist Focus Group and Survey

In their role, the County Reading Specialists were pivotal in addressing several of the major purposes of Reading First and their assessments inform the four pillars as well as four of the evaluation questions:

- Professional Development and Knowledge Transfer to Teachers: How effective was the professional development approach in helping teachers acquire knowledge and skills about phonemic awareness, phonics, fluency, vocabulary development, and reading comprehension, and transfer the knowledge and skills to their classroom instruction?
- Transfer of Knowledge in the Classroom: To what extent are teachers incorporating reading assessments into their classrooms and using the results of the assessments to change their instructional approaches and address students' learning needs?
- Capacity Building - Support System: How effective is the system of support for schools and districts to help all key stakeholders to contribute to the improvement of students' reading performance and sustain improved performance over time?
- Capacity Building – Leadership: To what extent has Reading First help develop instructional leadership in coaches, principals, and LEAS?

Additional effort was made to capture their attitudes and opinions, in this case specific to their own work and challenges. The questions went through several revisions in order to be able to describe, both quantitatively and qualitatively, their role and responsibilities in providing support and assisting teachers, schools and districts in the implementation of Reading First. The focus group and survey were administered on April 23, 2004. In general, the entire session, and in particular the focus group, may be characterized as cooperative and energetic, with everyone actively engaged in the group discussion. Seventeen specialists were present.

The purpose of the focus group was to examine the following: (1) the role and responsibilities of the county reading specialists; (2) the perceived changes and successes in the Reading First-schools in the past year and the role they played in the changes and successes; (3) the challenges of the role; and (4) the degree to which the training they received prepared them for the role as reading specialists.

Before the focus group discussion, the reading specialists were administered a short seven-item survey to obtain information on questions similar to those addressed in the focus group. The purpose here was to generate responses from individuals to see whether they differed from the group responses in order to isolate the group's influence as well as to provide a quantitative framework for presenting the discussion. Other items unique to the specialists were also included in the survey. There were 14 out of 16 of the specialists whom completed the survey, as two were absent from the group.

The survey and focus group facilitator was aided by two recorders, one to take notes of the proceedings and the other to record on a flip chart the topics and issues generated by the group for review. The focus group session lasted for approximately 40-45 minutes. For subsequent analysis, the evaluator also audiotaped the group session.

Site Visit Methodology – Observations and Interviews

The Site Visits, using classroom observations and a series of interviews, were designed to again provide insight to many areas using a different methodology as a way to continue to valid the overall results. The site visits cover the four pillars (professional development, leadership, assessment systems, and instruction) and provide additional insights on four of the research questions:

- Professional Development and Knowledge Transfer to Teachers: How effective was the professional development approach in helping teachers acquire knowledge and skills about phonemic awareness, phonics, fluency, vocabulary development, and reading comprehension, and transfer the knowledge and skills to their classroom instruction?
- Transfer of Knowledge in the Classroom: To what extent are teachers incorporating reading assessments into their classrooms and using the results of the assessments to change their instructional approaches and address students' learning needs?
- Capacity Building - Support System: How effective is the system of support for schools and districts to help all key stakeholders to contribute to the improvement of students' reading performance and sustain improved performance over time?
- Capacity Building – Leadership: To what extent has Reading First help develop instructional leadership in coaches, principals, and LEAS?

About a third of all Arizona Reading First schools were visited during the first year of school implementation. To select site visit schools, the evaluation team categorized all the schools into groups by size, district and geographic location and then randomly selected schools within each groups. The following schools were selected and visited:

- Andalucia
- Challenger
- Davidson
- Ironwood
- J.R. Davis
- Lafe Nelson
- La Pera
- Lowell
- Maxine O. Bush
- Menlo Park
- Orange Grove
- P.T. Coe
- Pendergast
- Phoenix Thomas J. Pappas
- Red Mesa
- Roosevelt
- Salida del Sol
- Shaw Butte
- Stanfield
- Summit View
- Valley Horizon
- Whittier
- William C. Jack

Schools were notified by the ADE in December 2003 that they had been selected for site visits; they received more detailed information about the visits from the site visit team in January 2004.

A team of evaluators shared in the work of the site visit. In most cases, a single evaluator visited each school, although a few schools were visited by a team of two evaluators. In order to ensure

common understandings of the instruments and to enhance reliability, a three-day training for all site visitors took place in February 2004.

Prior to each site visit, reading coaches and/or principals were contacted to make arrangements for the visit and were asked to complete a questionnaire with basic information about their reading program. For each site visit, schools were asked to schedule interviews with the principal and reading coach, as well as the assessment coordinator, if there was one in the school. In addition, the visits included a focus group with up to 12 members of the Reading Leadership Team and observations of three reading classrooms. Each of these activities is described in greater detail below.

Interviews
Interviews with principals, reading coaches, and assessment coordinators covered a similar range of topics: the roles of each, the work of the Reading Leadership Team, the type and perceived effectiveness of professional development, support from the state, perceptions of instructional change at the school, use of assessments, as well as challenges and successes in the first year of implementation. The coach interview was somewhat longer than the principal or assessment coordinator interview.

Interviews were not taped; instead, extensive notes were recorded and then summarized for each school. Consequently, the quotes provided in this report are not verbatim but do represent as closely as possible the actual wording of the respondents.

Focus Groups
In order to obtain the perspectives of teachers at Reading First schools, focus groups were held with members of the Reading Leadership Team. In schools with very large Reading Leadership Teams, evaluators asked to limit the size of the focus group to about 12 individuals, in order to better facilitate discussion. In cases where additional teachers wished to speak with evaluators, adjustments were made to include more teachers or to talk to additional teachers at other times.

Focus groups were held in all but one of the 23 visited schools. The average participation rate was six staff members plus the evaluator. Usually, the principal and reading coach were asked not to attend the focus group in order to help ensure that participants felt comfortable expressing all of their opinions.

Focus groups are most useful for obtaining information about the range of opinions, rather than a great deal of depth about any single topic. For this reason, the evaluators focused the discussion on the work of the Reading Leadership Team (RLT) in that school, and obtaining perspectives on these issues: professional development, core reading program and supplementary materials, use of assessments, instruction in the five essential components of reading, and the delivery of reading interventions.

Classroom Observations
In many Reading First schools, reading instruction occurs throughout the primary grades during a single 90-minute block of time during the school day. This meant evaluators only had a short period of time in which to see as much instruction as possible. For this reason, evaluators

decided to visit three classes, at different grade levels, for 20-30 minutes each, well aware that this information would provide only a "snapshot" of the instruction that occurred at the school.

Evaluators randomly selected classes, and when planning visits with the reading coach, asked the coach to see the particular classes selected. They also informed the coach that teachers had the right to request *not* to be observed and that in such circumstances a different class could be substituted (such substitutions were not common).

In total, site visitors conducted 74 classroom observations which were spread fairly evenly across the grades: 16 kindergarten classroom observations; 20 first-grade classroom observations; 19 second-grade classroom observations; and 19 third-grade classroom observations.

During the observations, the evaluators focused on the work of the teacher. For example, if the teacher was working with a group of five students and other students worked with a paraprofessional or on their own, in groups or individually, the observation focused on the small group work of the teacher. Paraprofessionals were not explicitly observed, although their presence in the classroom was noted. Evaluators took detailed notes in consecutive five-minute intervals, recording chronologically what the teacher did and how students responded. After the observation, evaluators used their notes to record the components being addressed in each five-minute block during the observation, and rated a series of items including the clarity of the portion of the lesson observed, the level of student engagement, the level of appropriate monitoring and feedback, as well as various physical characteristics of the classroom.

Because of some concerns about inter-rater reliability, described below, in the reporting of results, ratings of observed instruction and on-going assessment of learning were collapsed into two broad categories. Ratings between zero and two were collapsed into the category "occasionally or not at all," while ratings of three or four were put into the "yes, definitely." These broader categories then provided more reliable, if less nuanced, estimates of lesson clarity, teacher modeling, student engagement, student opportunities to practice, and teacher provision of clear and frequent feedback.

Teacher Interviews

Because the observations in classrooms were short, site visitors met with each of the observed teachers for a brief (10-15 minute) interview after the observation. This permitted the teacher to explain the broader context of the observation, including the rest of the lesson for the day, the goals for the current week and how those goals were determined. This was also a time to talk about teacher impressions of the core reading program, of changes in the school under Reading First, and of how well the teachers felt that the new programs were serving the needs of ELL students in the schools.

Validity and Reliability

The term "validity" in research is used to describe the degree to which the data being collected are an accurate measurement of the information desired. It is crucial to know that the observations record information that actually describes elements of instruction and in particular, that they describe elements of instruction that have a real impact on student achievement.

Reliability refers to the degree to which a tool measures the same thing in the same way. When multiple observers are in classrooms using numerical ratings to summarize some of the information about instruction, it is important to ensure that each observer rates the same lesson in the same way.

The creation of the observation protocol was a multi-step process designed to maximize the validity of the tool within the time and budget constraints of the evaluation. The designers began by reviewing recent literature on those elements of reading instruction that have been shown to be clearly linked to differences in student achievement (Foorman and Schatschneider 2003; Taylor et al., 2000; Snow et al. 1998). In addition, they were influenced by the principal walk-thru instrument presented by West-Ed trainers to Arizona Reading First principals, which identified key areas for principals to focus on when they paid shorter visits to reading classrooms.

This work highlighted a few key areas: subject of the lesson, clarity of the lesson, on-going monitoring and adjustment to student understanding, clear feedback to students, classroom environment, opportunities to practice, and student engagement. A team of reading evaluators compiled a first draft of an observation tool and used this to visit a non-Reading First school in Portland, Oregon. There, two or three evaluators visited the same classroom at the same time and then completed a rating form. After the visit, they carefully compared and discussed ratings, identifying items that were harder to achieve agreement on. Preliminary inter-rater reliability was 81.3 percent (within one-point of agreement). Problematic items were revised and rubrics were developed to better clarify the basis for making decisions about the ratings on each item.

Following the training of the site visit team for Arizona, trial observations were conducted at an Arizona Reading First school that had not been selected for a site visit this year. Two evaluators conducted observations of eight lessons and rated their observations independently (inter-rater reliability was 91.2 percent). Evaluators then debriefed and further clarified the remaining questions. Even so, after the actual site visits, ratings of different site visitors were compared, and some evaluators appeared to rate consistently lower or higher than others. It is difficult to know whether the differences reflect true differences in the schools or differences in site visitor rating. In order not to be excessive weight on the difference between, for example, a "1" and a "2" rating, low (0-2 point) and high (3-4 point) ratings were collapsed for the analyses presented here.

In addition to recording ratings, evaluators also logged what was happening in the classroom, and these notes were used to provide the qualitative examples in the text.

District and State Interviews

The interviews were conducted to specifically address the capacity building of leadership and the support system, but also touched on other areas. Thus, questions touched on the four pillars (professional development, leadership, assessment systems, and instruction) and four of the research questions:

- Professional Development and Knowledge Transfer to Teachers: How effective was the professional development approach in helping teachers acquire knowledge and skills

about phonemic awareness, phonics, fluency, vocabulary development, and reading comprehension, and transfer the knowledge and skills to their classroom instruction?

- Transfer of Knowledge in the Classroom: To what extent are teachers incorporating reading assessments into their classrooms and using the results of the assessments to change their instructional approaches and address students' learning needs?
- Capacity Building - Support System: How effective is the system of support for schools and districts to help all key stakeholders to contribute to the improvement of students' reading performance and sustain improved performance over time?
- Capacity Building – Leadership: To what extent has Reading First help develop instructional leadership in coaches, principals, and LEAS?

District Interviews

Interviews were conducted with one or more representatives from 13 of the 19 visited districts. Interviewees included staff from various district departments, including curriculum and instruction, budget and finance offices, and the superintendent. These representatives were asked questions about the district's role in Reading First, how the grant fit with other work in the district, grant implementation, and district professional development and technical assistance to Reading First and non-Reading First schools.

State Interview

The Arizona Department of Education's (ADE) vision of Reading First was first articulated in Arizona's Reading First Plan. In order to deepen evaluators' understanding of this vision, and to clarify any modifications or additions to the original plan, evaluators interviewed the statewide Director of Arizona Reading First in June 2004. The one and one-half hour phone interview included specific questions about the state's vision and direction in the following areas: role of the principal, Reading Leadership Team, and district in grant implementation; structure and content of the 90-minute reading block; definition of "fidelity"; supplemental and intervention programs; use of assessment data; role of the ADE in providing professional development; and challenges and successes in the first year of implementation.

CHAPTER 4
STUDENT ASSESSMENTS: DIBELS BY GRADE LEVEL

Highlights

Students were assessed with the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). In kindergarten and first grade, the assessment covers several areas of reading skills, while in second and third grade, the assessment tests oral reading fluency. In all grades, student scores can be translated into overall instructional recommendations of "benchmark" (students on target to read at grade level), "strategic" (student in need of strategic interventions in order to read at grade level) and "intensive" (students in need of intensive intervention).

- All of the Reading First grade level groups showed positive change.

Reading First students

- The percent of students at "benchmark" **increased** for each of the grade levels over the percent at benchmark at the beginning of the year.

- Kindergarten difference in the benchmark category from the beginning to end of the year was the most dramatic with a 44 percent increase.

- While first graders saw a small increase in those at benchmark from the beginning to end of year (7.4 percent increase), the second and third graders change in percent at benchmark was almost flat with a 3.5 and 2.7 increase respectively.

- The percent in the benchmark category at the end of the year was highest in kindergarten at over 50 percent, with the first graders at 41 percent and only about one-third of second and third graders at benchmark at the end of the year
(K = 53.1%, first = 40.9%, second = 32.2%, and third = 30.5%).

- Kindergarten students began the year at a very low level with average scores of less than 5 on the ISF and LNF measures and 50 percent or more categorized "at risk;" by the end of the year, only 30 percent of students were "at risk" on the LNF and 45 to 62 percent of students were categorized at "low risk" on the LNF, PSF and NWF.

- In first grade, the NWF average scores increased by over 40 points from beginning to end of the year; the NWF category had a substantial drop (34 points) in the percent "at risk" and a substantial increase (28 points) in those at "low risk" at the end of the year.

- In first grade, the PSF scores changed by 28 points from beginning to end; at the end, only 1.4 percent of students were "at risk" with 88.5 percent at "low risk."

- The second grade ORF average score increased by almost 30 points from the beginning to the end of the year; however although the percent at "low risk" increased by several points also did the percent "at risk" increase by several points.

- The third grade ORF average score increased by 32 points; the percent "at risk" dropped 10 points and the percent at "low risk" increased just 3 points.

- The DIBELS measure was used to compare the performance of Reading First schools to a group of comparison schools that already administered the DIBELS measure as well as to a second comparison group that did not and that had the DIBELS test administered to a sample of students at the end of the year; this group is referred to as "Post only" group (see Figures 4.1 and 4.2).

- Reading First kindergarteners improved more than the comparison group. Although Reading First students started out lower, they ended up with a slightly higher percent at end of year benchmark; the Reading First and comparison group ended up with a higher percent at benchmark than the Post only group.

- First, second and third grade Reading First students improved, whereas the comparison groups changed in a negative direction with the percent in benchmark at the end of the year lower than the percent at the beginning of the year and also lower than the Reading First group at the end of year. The Reading First group had a higher percent at end in benchmark than either the comparison groups or the Post only groups.

Figure 4.1

Percentage of students at Benchmark - Beginning of Year

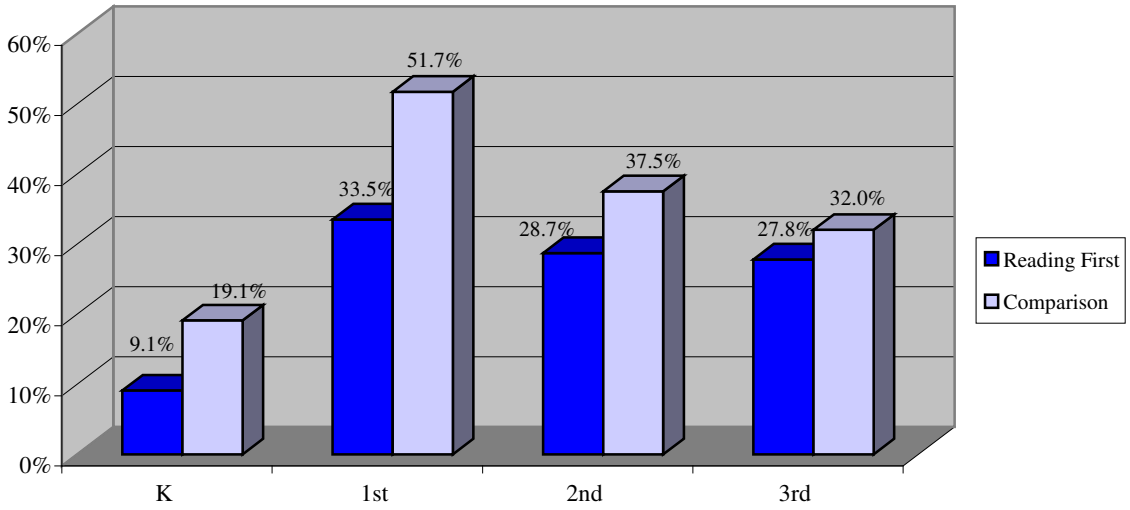
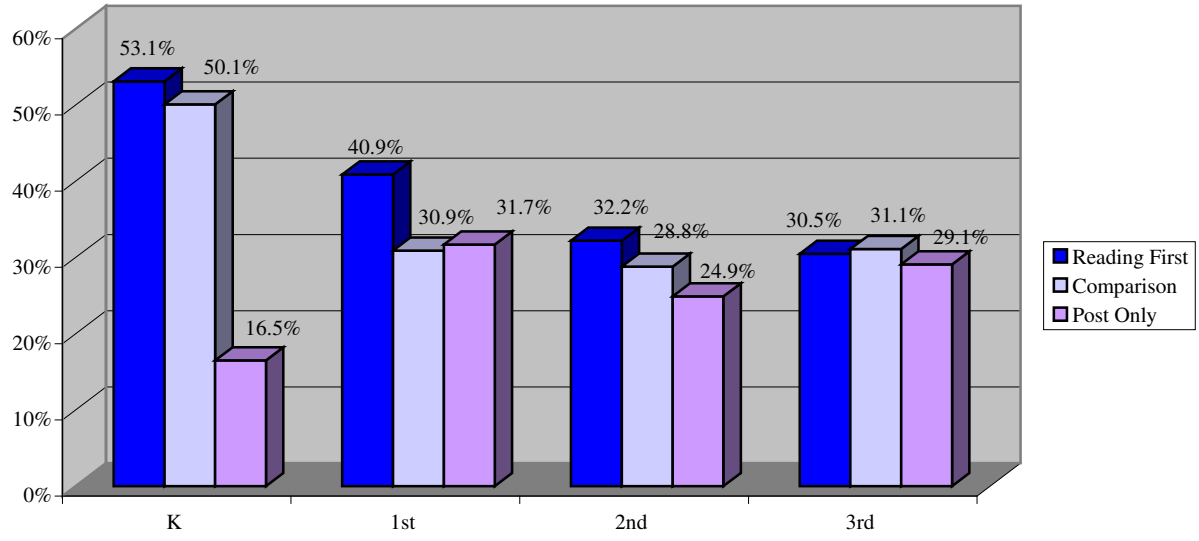


Figure 4.2

Percentage of students at Benchmark - End of Year



DIBELS Test Measures and Categorization

The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) are a set of standardized, individually administered measures of early literacy development. They are designed to be short (one minute) fluency measures used to regularly monitor the development of pre-reading and early reading skills. The measures were developed upon the essential early literacy domains discussed in both the National Reading Panel (2000) and National Research Council (1998) reports to assess student development of phonological awareness, alphabetic understanding, and automaticity and fluency with the code. Each measure has been thoroughly researched and demonstrated to be reliable and valid indicators of early literacy development and predictive of later reading proficiency to aid in the early identification of students who are not progressing as expected. When used as recommended, the results can be used to evaluate individual student development as well as provide grade-level feedback toward validated instructional objectives (http://reading.uoregon.edu).

The DIBELS measures were specifically designed to assess 3 of the 5 Big Ideas of early literacy: Phonological Awareness, Alphabetic Principle, and Fluency with Connected Text. The measures are linked to one another, both psychometrically and theoretically, and have been found to be predictive of later reading proficiency.

- Measures of Phonological Awareness:
    - Initial Sounds Fluency (ISF): Assesses a child's skill to identify and produce the initial sound of a given word.
        - The ISF is one of two DIBELS measures of phonemic awareness and can be administered to students who cannot yet read. Students are simply asked to point to the picture that starts with a particular sound (the examiner provides the name for each picture). The score reports the number of correct pictures identified.
    - Phonemic Segmentation Fluency (PSF): Assesses a child's skill to produce the individual sounds within a given word.
        - The PSF is a measure of phonological awareness assessing student ability to break a word into individual sounds, or phonemes. The score on this assessment consists of the number of correct phonemes produced in one minute. The ultimate target is for students to be able to correctly segment at least 35 phonemes (18 by the middle of kindergarten).

- Measure of Alphabetic Principle:
    - Nonsense Word Fluency (NWF): Assesses a child's knowledge of letter-sound correspondences as well their ability to blend letters together to form unfamiliar "nonsense" (e.g., fik, lig, etc.) words.
        - The NWF measures student knowledge of letter-sound correspondence to sound out “pretend” words they have never encountered. The score is the number of letter-sounds produced correctly in one minute, and the ultimate target is a score of 50 by the middle of first grade.

- Measure of Fluency with Connected Text:
    - Oral Reading Fluency (ORF): Assesses a child's skill of reading connected text in grade-level material word.
        - The Oral Reading Fluency (ORF) is a measure of reading fluency (rate and accuracy) with grade-level text. The students are asked to read a passage aloud for one minute. It is administered to first-grade students twice a year and to second- and third-grade students, three times a year. The score is the number of words read minus the number of errors, or the words correct per minute (wcpm).

- Phonics
    - Letter Naming Fluency (LNF): This particular assessment provides a measure of risk. There is no benchmark goal provided because it has no correspondence with one of the big ideas of early literacy skills. In addition, it does not seem to be fundamental for the achievement of reading outcomes.
        - The LNF measure asks students to identify as many randomly presented upper- and lowercase letters as possible in one minute. While it is not an actual phonics skill, it is highly predictive of future reading success. It is administered to kindergarten students three times a year.

- Word Use Fluency (WUF): At this time there is no benchmark goal provided because further research is still needed in order to establish its connection with the big ideas of early literacy skills.

- Retell Fluency (RTF): is a measure that assesses comprehension, the ability to extract meaning from text. It is intended to offer a comprehension check for the ORF assessment. The purpose of the RTF measure is to (a) prevent inadvertently learning or practicing a misrule, (b) identify children whose comprehension is not consistent with their fluency, (c) provide an explicit linkage to the core components in the NRP report, and (d) increase the face validity of the ORF.

These measures link together to form an assessment system of early literacy development that allows educators to readily and reliably determine student progress. Students are assessed with these various measures that are then combined to categorize students based on their scores. Students are either "at risk," "some risk," or "low risk." This categorization becomes the basis for the overall reading recommendation which classifies students as either "intensive," strategic" or "benchmark" in terms of their overall performance. These categories and scores can help teachers plan interventions specific to the needs of each student. For assessment, it aids in viewing how students are grouped and how they progress.

Schools received Arizona Reading First grants in part because they served a large proportion of students who read below grade level. A central purpose of the grant was to provide schools with the structures and tools to move low-performing students up to grade level in reading. Using the DIBELS, it was possible to track the movement of students out of the "intensive" (lowest) or "strategic" (middle) groups into the "benchmark" (highest) group. It was also possible to

determine whether students who started the year off at benchmark made enough gains over the year to remain at benchmark at the end of the year in the spring.

Calculation of Overall Instructional Support Recommendations

The DIBELS system classifies the categories of scores into a measure called the Instructional Support Recommendation. Thus depending upon how a student is categorized on each of the measures, a condensed measure that combines all the scores and categories allows for school staff to classify the overall level of the student as "intensive" (students in need of intensive interventions to reach grade level reading), "strategic" (students in need of targeted support with specific skills) or "benchmark" (students who are on track to read at grade level). These instructional recommendations offer a good sign of the overall status of student reading skills at a particular point in time.

For the second and third grade, instructional support recommendations came directly from their performance on the ORF section of the DIBELS. For the earlier grades, instructional recommendations were based on a combination of student performance on multiple measures. For kindergarten, it was the ISF, LNF, PSF and NWF scores in combination that determine whether a student, overall, requires "intensive," "strategic" or "benchmark" level instruction. For the first grade, the PSF and NWF scores are considered together with the ORF in order to determine instructional recommendations. Evaluators followed the guidelines of the DIBELS developers in order to combine the scores and determine instructional recommendations.

Not all measures are administered to all students at each testing period; instead, only those measures are administered that apply to skills students should be mastering at a particular period. Thus for example, Phoneme Segmentation Fluency is assessed beginning in the middle of kindergarten and continuing through the end of first grade, while Oral Reading Fluency in not assessed until the winter of first grade. Table 4.1 indicates which measure is administered to each grade level at each assessment period.

Table 4.1
Scheduled Administration of DIBELS Assessment Measures

Measure	Fall	Winter	Spring
Initial Sound Fluency (ISF)	K	K	--
Letter Naming Fluency (LNF)	K, 1	K	K
Phoneme Segmentation Fluency (PSF)	1	K, 1	K, 1
Nonsense Word Fluency (NWF)	1	K, 1	K, 1
Oral Reading Fluency (ORF)	2, 3	1, 2, 3	1, 2, 3

Findings: Kindergarten

There were 5,580 matched DIBELS assessment scores for RF kindergarten students during the fall and spring testing periods.

The performance of Reading First schools was compared to a group of comparison schools that already administered the DIBELS measure as well as to a second comparison group that did not and that had the DIBELS test administered to a sample of students at the end of the year; this group is referred to as "Post only" group

DIBELS Measures
At the beginning of the year, kindergarten students were assessed on the ISF (phonological awareness) and LNF (phonics) measures. At the end of the year, they were assessed on the PSF (phonological awareness), LNF (phonics), and NWF (alphabetic principle) measures. The phonological awareness measure, in this case the PSF, is the most important predictor of future student success. Per the DIBELS system, the raw scores were coded into the "at risk," "some risk" and "low risk" categories.

The results indicated that overall the RF kindergarten students improved during the year. For example, on the LNF measure, students' mean beginning score was 4.58 compared to an end LNF measure of 37.69. On the two measures at the beginning of the year, approximately 50 percent or more of RF students were in the "at risk" categories, with 12 to 30 percent "at risk" at the end of the year on the three measures (Table 4.2). Conversely, 25 percent or less started the year at "low risk" with 45 to 62 percent achieving "low risk" by the end of year.

Specifically, at the beginning of the year, 61.5 percent of students were "at risk" on the LNF and this dropped to 30.3 percent at the end of the year; correspondingly, only 19.1 percent were "low risk" in the beginning with 45.4 percent rising to this category by the end of the year. At the end of the year, the PSF measure had 62.1 percent assessed as "low risk" while the NWF had 57.8 percent as "low risk."

Table 4.2
Reading First - Kindergarten Category Measures

READING FIRST	% Beginning of Year ISF	% Beginning of Year LNF	% End of Year LNF	% End of Year PSF	% End of Year NWF
Mean Score	4.98	4.58	37.69	36.55	29.06
At Risk	48.1 N = 2681	61.5 N = 3433	30.3 N = 1693	12.8 N= 715	22.7 N = 1264
Some Risk	27.0 N = 1504	19.4 N = 1081	24.3 N = 1353	25.1 N= 1398	19.6 N = 1092
Low Risk	25.0 N = 1394	19.1 N = 1066	45.4 N = 2533	62.1 N = 3464	57.8 N = 3222
Total	100.0 N = 5579	100.0 N = 5580	100.0 N = 5579	100.0 N = 5577	100.0 N = 5578

There are some slight differences by category between the RF and comparison group (Table 4.3). The larger differences occurred when these groups were compared to the Post only group that ended the year with 30 to 40 percent more students "at risk" (Table 4.4). As seen on Table 4.3, the average beginning scores on the ISF and LNF were slightly higher for the comparison group than the RF group. At the end of the year, the RF group had caught up to the comparison group as average scores at the end were slightly higher for the RF group than the comparison group on the PSF and NWF and within one point on the LNF. The average end of year scores for the Post Only comparison group were markedly lower (13 to17 points) on all measures compared with either the RF or comparison group (Table 4.4).

Table 4.3
Comparison Schools Category Measures

COMPARISON SCHOOLS	% Beginning of Year ISF	% Beginning of Year LNF	% End of Year LNF	% End of Year PSF	% End of Year NWF
Mean Score	5.39	10.09	38.42	32.85	26.01
At Risk	47.6 N = 418	37.8 N = 332	29.2 N = 257	18.8 N = 165	30.9 N = 272
Some Risk	23.0 N = 202	22.0 N = 193	24.9 N = 219	31.4 N = 276	17.0 N = 150
Low Risk	29.5 N = 259	40.3 N = 354	45.9 N = 404	49.9 N = 439	52.0 N = 458
Total	100.0 N = 879	100.0 N = 879	100.0 N = 880	100.0 N = 880	100.0 N = 880

Table 4.4
Post Only Schools Category Measures

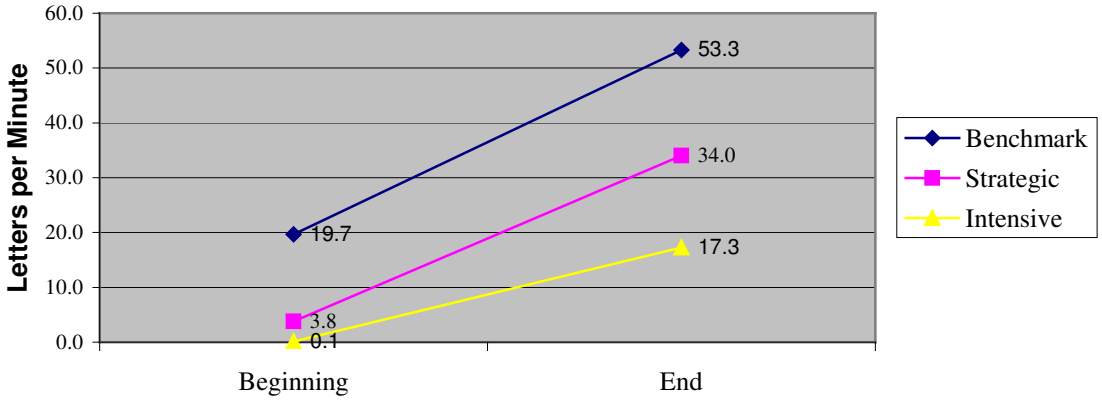
POST ONLY SCHOOLS	% End of Year LNF	% End of Year PSF	% End of Year NWF
Mean Score	24.13	15.53	12.63
At Risk	59.9 N = 151	49.8 N = 125	62.8 N = 157
Some Risk	19.8 N = 50	33.5 N = 84	20.8 N = 52
Low Risk	20.2 N = 51	16.7 N = 42	16.4 N = 41
Total	100.0 N = 252	100.0 N = 251	100.0 N = 250

Despite starting with a few students at benchmark, by the end of the year, Reading First schools had seen a dramatic increase in the percentage of students at benchmark.

Figure 4.3 shows the mean score for kindergarten Reading First students in each of the categories on the LNF measure. Therefore, for the "benchmark" students, the mean score at the beginning of year was 19.7 and at the end of the year it was 53.3, a growth of 33.6 points.

Figure 4.3

Kindergarten Reading First Letter Naming Fluency (LNF)
Growth Chart



DIBELS Instructional Support Recommendation

The DIBELS system classifies the categories of scores into a measure called the Instructional Support Recommendation. Thus depending upon how a student is categorized on each of the measures, a condensed measure that combines all the scores and categories allows for school staff to classify the overall level of the student as "intensive" (students in need of intensive interventions to reach grade level reading), "strategic" (students in need of targeted support with specific skills) or "benchmark" (students who are on track to read at grade level). These instructional recommendations offer a good sign of the overall status of student reading skills at a particular point in time.

At the beginning of the year, 58.8 percent of the RF students were at "intensive," compared to 44.0 percent of the comparison group. Only 9.1 percent of RF students and 19.1 percent of the comparison group were at "benchmark" at the beginning. By the end of the year, the RF and comparison group were similar with 53.1 and 50.1 percent respectively classified as benchmark (see Table 4.5). The end of year comparison between groups, then, showed that these two groups were very similar in terms of the percentages in each of the three categories. Again, the difference was between these groups and post-test only group; only 16.5 percent of the Posttest only comparison group reached benchmark and 63.1 percent were at the "intensive" level at the end of kindergarten. What is unknown is whether these Post only scores are "typical" for the "normal" Arizona kindergarten that has traditionally emphasized socialization over reading skill development.

The RF kindergarten students improved by the posttest at the end of the year and these students also started out lower in the pretest than the comparison students. In the posttest, the RF students caught up with their counterparts in the pretest-posttest comparison schools and, in fact, did slightly better in the spring than the comparison group.

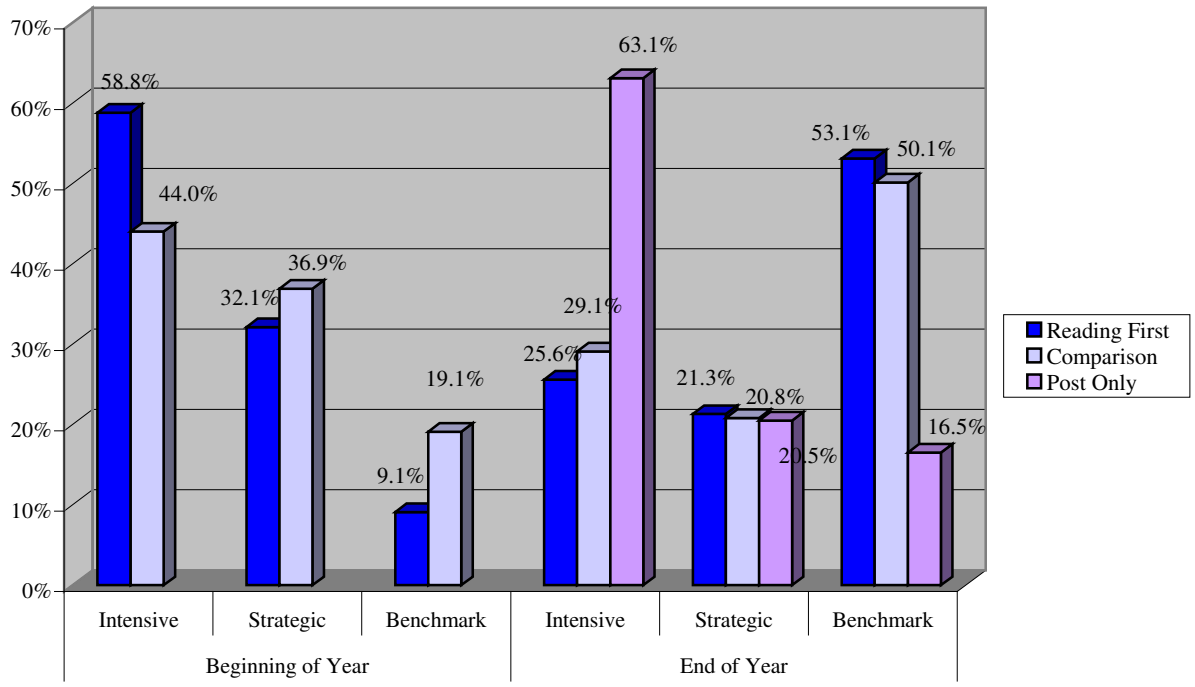
What is known is that of the eight comparison schools, four were in Reading First LEAs and 5 had science-based reading core curricula in place. For the eight Post only schools, two were in RF LEAs and 3 had science-based reading core curricula in place.

Table 4.5
Kindergarten Instructional Support Recommendation

		Kindergarten Instructional Recommendation Beginning of Year				Kindergarten Instructional Recommendation End of Year			
		Intensive	Strategic	Benchmark	Total	Intensive	Strategic	Benchmark	Total
Reading First	N	3281	1792	506	5579	1429	1185	2962	5576
	%	58.8%	32.1%	9.1%	100.0%	25.6%	21.3%	53.1%	100.0%
Comparison	N	386	324	168	878	256	183	441	880
	%	44.0%	36.9%	19.1%	100.0%	29.1%	20.8%	50.1%	100.0%
Post Only	N					157	51	41	249
	%					63.1%	20.5%	16.5%	100.0%
Total	N	3667	2116	674	6457	1842	1419	3444	6705
	%	56.8%	32.8%	10.4%	100.0%	27.5%	21.2%	51.4%	100.0%

Figure 4.4

Kindergarten Instructional Support Recommendation



There was only a 3.1 percent difference between the RF and comparison groups of those who ended up at "benchmark" whereas there were 36.6 and 33.6 percent differences of RF and the comparison group to the end of year percent at "benchmark" for the Post only group (see Table 4.6).

Table 4.6
Kindergarten Difference Comparison End of Year Benchmark

Comparison End of Year Benchmark	% Difference
RF to Comparison group	3.1%
RF to Post Only	36.6%
Comparison to Post Only	33.6%

Ethnicity

Table 4.7 shows the change in the benchmark category from the beginning to end of the year by ethnicity. According to the data, of the total 506 students at benchmark at the beginning of the year, 183 (36.2%) were Hispanic/Latino students and 169 (33.4%) were White, not Hispanic/Latino students. By the end of the year, Hispanic/Latino students were 53.6 percent of the2962 students at benchmark, while White, not Hispanic/Latino students were 14.9 percent of the 2962 students at benchmark.

Table 4.7
Beginning & End of the Year Instructional Support Recommendation
% at Benchmark by Ethnicity

		Kindergarten – Beginning of the Year ISR	Kindergarten - End of the Year ISR
		Benchmark	Benchmark
Not Reported	Count	78	657
	%	15.4%	22.2%
American Indian	Count	33	142
	%	6.5%	4.8%
Asian	Count	12	36
	%	2.4%	1.2%
Black/African-American, not Hispanic/Latino	Count	29	92
	%	5.7%	3.1%
Hispanic/Latino	Count	183	1587
	%	36.2%	53.6%
Native Hawaiian/Pacific Islander	Count	2	6
	%	0.4%	0.2%
Other	Count	0	1
	%	.0%	0.03%
White, not Hispanic/Latino	Count	169	441
	%	33.4%	14.9%
Total	Count	506	2962
	%	100%	100%

Free & Reduced Lunch

Free and reduced lunch is an indicator of student poverty. Overall, a very large percentage of Reading First students are eligible for free and reduced lunch. Table 4.8 shows the change in the benchmark category from the beginning to end of the year according to those students who qualify for free and reduced lunch. The data showed that of the total 506 students at benchmark at the beginning of the year, 163 (32.2%) were eligible for free and reduced lunch. By the end of the year, students eligible for free and reduced lunch were 36 percent of the 2962 students at benchmark.

Table 4.8
Beginning & End of the Year Instructional Support Recommendation
% at Benchmark by Free & Reduced Lunch

		Kindergarten – Beginning of the Year ISR	Kindergarten - End of the Year ISR
		Benchmark	Benchmark
Not Reported	Count	259	1561
	%	51.2%	52.7%
Eligible	Count	163	1067
	%	32.2%	36.0%
Not Eligible	Count	84	334
	%	16.6%	11.3%
Total	Count	506	2962
	%	100%	100%

Special Education

Table 4.9 shows the percentages in the benchmark category from the beginning to end of the year according to those students who qualify for special education. According to the data, of the total 506 students at benchmark at the beginning of the year, 12 (2.4%) were eligible for special education. By the end of the year, students eligible for special education were 3.3 percent of the 2962 students at benchmark.

Table 4.9
Beginning & End of the Year Instructional Support Recommendation
% at Benchmark by Special Education

		Kindergarten – Beginning of the Year ISR	Kindergarten - End of the Year ISR
		Benchmark	Benchmark
Not Reported	Count	294	1681
	%	58.1%	56.8%
Eligible	Count	12	99
	%	2.4%	3.3%
Not Eligible	Count	200	1182
	%	39.5%	39.9%
Total	Count	506	2962
	%	100%	100%

English Language Learners (ELL)
Table 4.10 shows the change in the benchmark category from the beginning to end of the year according to those students who qualify for ELL services. According to the data, of the total 506 students at benchmark at the beginning of the year, 43 (8.5%) were ELL students. By the end of the year, ELL students were 26.6 percent of the2962 students at benchmark.

Table 4.10
Beginning & End of the Year Instructional Support Recommendation
% at Benchmark by ELL

		Kindergarten – Beginning of the Year ISR	Kindergarten - End of the Year ISR
		Benchmark	Benchmark
ELL % at Benchmark	Count	43	789
	%	8.5%	26.6%
Non-ELL Students % at Benchmark	Count	463	2173
	%	91.5%	73.4%
Total % at Benchmark	Count	506	2962
	%	100%	100%

Findings: First Grade

There were 5,749 matched DIBELS assessment scores for first grade during the fall and spring testing periods.

DIBELS Measures
In the fall, first grade students were assessed on the LNF (phonics), PSF (phonological awareness), and NWF (alphabetic principle) measures. In the spring, they were assessed again on the PSF (phonological awareness) and NWF (alphabetic principle connected to fluency) in addition to the ORF (fluency with connected text). The fluency with connected text measure, in this case the ORF, is the most important predictor of future student success. In this discussion, growth in shown for the two measures, PSF and NWF, that are assessed at both the beginning and end of the year. However, since the desired outcome is truly the ORF measure, its relationship to the NWF beginning measure is discussed in-depth in the next chapter of this report.

The results indicated that overall the RF first grade students improved during the year. On the PSF and NWF measures, 35 to 43 percent of RF students respectively were in the "at risk" categories at the beginning of the year, with 1 to 9 percent "at risk" at the end of the year. Specifically, on the PSF measure, students' mean beginning score was 20.73 compared to an end PSF measure of 48.72. On the ORF measure at the end of year, approximately 30 percent were "at risk." Conversely, 33 percent or less started the year at "low risk" with 40 to 89 percent achieving "low risk" by the end of year (Table 4.11).

Table 4.11
Reading First – First Grade Category Measures

READING FIRST	% Beginning of Year LNF	% Beginning of Year PSF	% Beginning of Year NWF	% End of Year NWF	% End of Year PSF	% End of Year ORF
Mean Score	28.26	20.73	18.6	61.75	48.72	39.02
At Risk	42.2 N = 2426	35.5 N = 2042	43.3 N = 2492	8.9 N = 513	1.4 N = 78	29.7 N = 1708
Some Risk	27.8 N = 1596	39.6 N = 2275	23.2 N = 1336	29.9 N = 1721	10.1 N = 583	29.4 N = 1688
Low Risk	30.0 N = 1727	24.9 N = 1429	33.4 N = 1921	61.1 N = 3514	88.5 N = 5088	40.9 N = 2352
Total	100.0 N = 5749	99.9 N = 5746	100.0 N = 5749	100.0 N = 5748	100.0 N = 5749	100.0 N = 5748

In first grade, the average beginning scores on the measures were higher for the comparison group than the RF group (6-9 points); the comparison group started at a slightly higher level. At the beginning of the year approximately 12 to 17 percent less students in the comparison group were in the "at risk" category than RF students, whereas at the end of the year the comparison

group had approximately 2 to 10 percent more students in the "at risk" category than RF (Table 4.12). At the end, the RF average scores were higher than the comparison group (NWF and ORF) and less than one-quarter point lower on the PSF. The average scores for the Post only comparison group were lower (8 to 20 points) on all measures than the RF group (Table 4.13). The Post only group ended the year with 4 to 20 percent more students "at risk" than the other two groups.

It is clear from Table 4.11above that Reading First schools saw a decline in the percentage of students in the "at risk" category on the phonemic awareness and decoding measures (PSF and NWF). As Table 4-12 illustrates, comparison schools saw a similar decline. For both measures, the post-only schools had a higher percentage of students "at risk" in phonemic awareness and decoding (Table 4.13).

While the students in RF and comparison schools looked similar on the phonemic awareness and decoding measures, performance on the oral reading fluency component looked different. RF schools had a lower percentage of students "at risk" in the ORF measure than the comparison and post-only schools did.

Table 4.12
Comparison Schools Category Measures

COMPARISON SCHOOLS	% Beginning of Year LNF	% Beginning of Year PSF	% Beginning of Year NWF	% End of Year NWF	% End of Year PSF	% End of Year ORF
Mean Score	36.45	29.88	24.02	52.62	48.89	32.56
At Risk	24.6 N = 231	22.5 N = 211	31.1 N = 292	17.7 N = 166	3.2 N = 30	40.3 N = 378
Some Risk	23.6 N = 222	30.8 N = 289	22.6 N = 212	37.2 N = 349	12.5 N = 117	28.9 N = 271
Low Risk	51.8 N = 486	46.6 N = 438	46.2 N = 434	45.2 N = 424	84.3 N = 792	30.9 N = 290
Total	100.0 N = 939	99.9 N = 938	99.9 N = 938	100.0 N = 939	100.0 N = 939	100.0 N = 939

Table 4.13
Post Only Schools Category Measures

POST ONLY SCHOOLS	% End of Year NWF	% End of Year PSF	% End of Year ORF
Mean Score	42.05	32.99	31.97
At Risk	28.1 N = 54	7.8 N = 15	41.7 N = 80
Some Risk	46.4 N = 89	39.1 N = 75	26.0 N = 50
Low Risk	25.5 N = 49	52.1 N = 100	31.3 N = 60
Total	100.0 N = 192	99.0 N = 190	99.0 N = 190

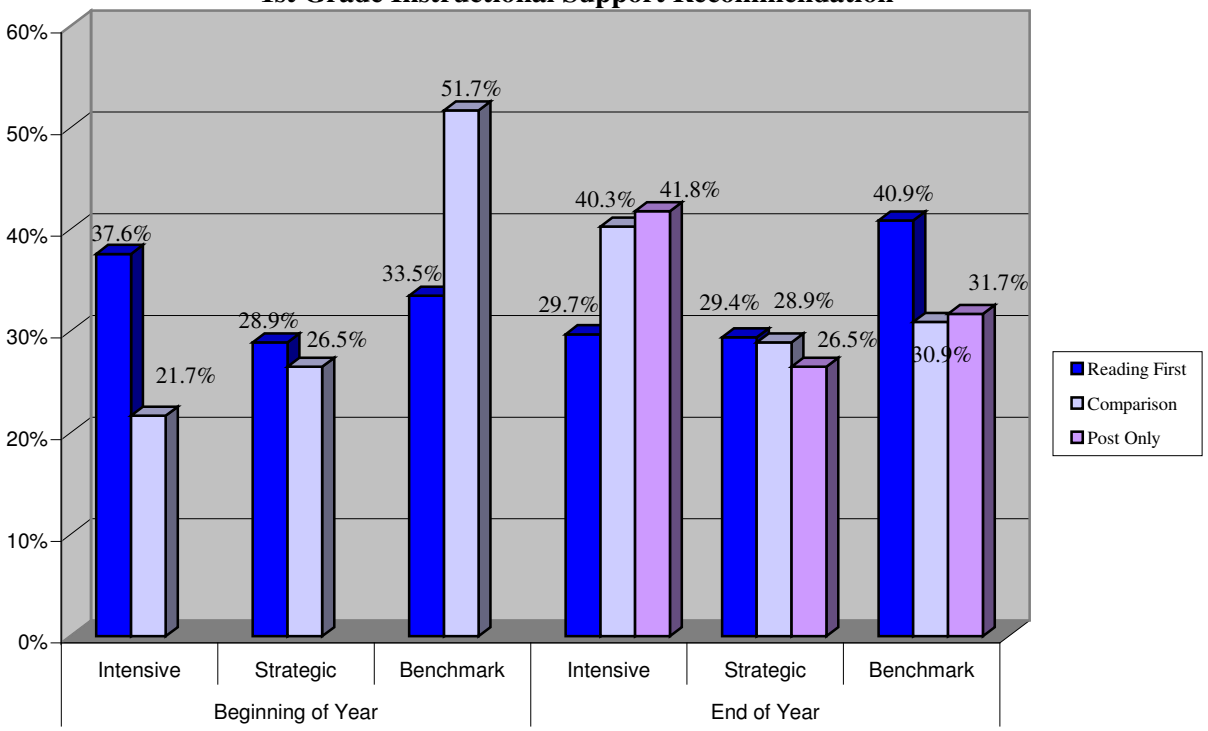
DIBELS Instructional Support Recommendation
The first grade RF students improved slightly by the posttest, the percent at benchmark went from 33.5 percent at the beginning to 40.9 percent at the end. Although this is not a large increase, it is a positive finding for the overall reading program. In contrast, the comparison group showed a decrease in the percent in the benchmark category. The comparison group began the year with 51.7 percent in the benchmark group and ended the year with on 30.9 percent in the benchmark group, whereas the RF group ended with 40.9 percent at benchmark. The first graders in the comparison group did not achieve noticeable growth in reading based upon percentages in the benchmark category during first grade, while the RF students made some gains. The Post Only group ended the year with their final benchmark percent at 31.7, similar to the comparison group at only 30.9 percent at benchmark (Table 4.14).

Table 4.14
First Grade Instructional Support Recommendation

		First Grade Instructional Recommendation Beginning of Year				First Grade Instructional Recommendation End of Year			
		Intensive	Strategic	Benchmark	Total	Intensive	Strategic	Benchmark	Total
Reading First	N	2163	1658	1925	5746	1708	1688	2351	5747
	%	37.6%	28.9%	33.5%	100.0%	29.7%	29.4%	40.9%	100.0%
Comparison	N	204	249	485	938	378	271	290	939
	%	21.7%	26.5%	51.7%	100.0%	40.3%	28.9%	30.9%	100.0%
Post Only	N					79	50	60	189
	%					41.8%	26.5%	31.7%	100.0%
Total	N	2367	1907	2410	6684	2165	2009	2701	6875
	%	35.4%	28.5%	36.1%	100.0%	31.5%	29.2%	39.3%	100.0%

Figure 4.5

1st Grade Instructional Support Recommendation



The RF group varied by approximately a 10 percent difference at "benchmark" between the comparison group (10%) and the Post Only group (9.2%). The comparison group and the Post Only group had very similar percentages of students in the "benchmark" category at the end of the year, differing only by –0.8 percent (see Table 4.15).

Table 4.15
First Grade Change Comparison End of Year Benchmark

Comparison End of Year Benchmark	% change
RF to Comparison group	10.0
RF to Post Only	9.2
Comparison to Post Only	-0.8

Ethnicity

Table 4.16 shows the change in the benchmark category from the beginning to end of the year by ethnicity. According to the data, of the total 1,925 students at benchmark at the beginning of the year, 1,162 (60.4%) were Hispanic/Latino students and 397 (20.6%) were White, not Hispanic/Latino. By the end of the year, Hispanic/Latino students were 62.3 percent of the 2351 students at benchmark, while White, not Hispanic/Latino students were 19 percent of the 2351 students at benchmark.

Table 4.16
Beginning & End of the Year Instructional Support Recommendation
% at Benchmark by Ethnicity

		First Grade End of Year ISR	First Grade End of Year ISR
		Benchmark	Benchmark
Not Reported	Count	129	191
	%	6.7%	8.1%
American Indian	Count	128	128
	%	6.6%	5.4%
Asian	Count	33	38
	%	1.7%	1.6%
Black/African-American, not Hispanic/Latino	Count	72	79
	%	3.7%	3.4%
Hispanic/Latino	Count	1162	1464
	%	60.4%	62.3%
Native Hawaiian/Pacific Islander	Count	4	5
	%	0.2%	0.2%
White, not Hispanic/Latino	Count	397	446
	%	20.6%	19.0%
Total	Count	1925	2351
	%	100%	100%

Free & Reduced Lunch
Table 4.17 shows the change in the benchmark category from the beginning to end of the year according to those students who qualify for free and reduced lunch. The data showed that of the total 1925 students at benchmark at the beginning of the year, 857 (44.5%) were eligible for free and reduced lunch. By the end of the year, students eligible for free and reduced lunch were 47.7 percent of the 2352 students at benchmark.

Table 4.17
Beginning & End of the Year Instructional Support Recommendation
% at Benchmark by Free & Reduced Lunch

		First Grade End of Year ISR	1st Grade End of Year ISR
		Benchmark	Benchmark
Not Reported	Count	845	971
	%	43.9%	41.3%
Eligible	Count	857	1122
	%	44.5%	47.7%
Not Eligible	Count	223	259
	%	11.6%	11.0%
Total	Count	1925	2352
	%	100%	100%

Special Education
Table 4.18 shows the change in the benchmark category from the beginning to end of the year according to those students who qualify for special education. According to the data, of the total 1925 students at benchmark at the beginning of the year, 79 (4.1%) were eligible for special education. By the end of the year, students eligible for special education were 4.7 percent of the 2352 students at benchmark.

Table 4.18
Beginning & End of the Year Instructional Support Recommendation
% at Benchmark by Special Education

		First Grade Beginning of Year ISR	First Grade End of Year ISR
		Benchmark	Benchmark
Not Reported	Count	974	1193
	%	50.6%	50.7%
Eligible	Count	79	110
	%	4.1%	4.7%
Not Eligible	Count	872	1049
	%	45.3%	44.6%
Total	Count	1925	2352
	%	100%	100%

English Language Learners (ELL)

Table 4.19 shows the change in the benchmark category from the beginning to end of the year according to those students who qualify for ELL services. According to the data, of the total 1925 students at benchmark at the beginning of the year, 539 (28%) were ELL students. By the end of the year, ELL students were 32.8 percent of the students at benchmark.

Table 4.19
Beginning & End of the Year Instructional Support Recommendation
% at Benchmark by ELL

		First Grade Beginning of Year ISR	First Grade End of Year ISR
		Benchmark	**Benchmark**
ELL % at Benchmark	Count	539	771
	%	28%	32.8%
Non-ELL Students % at Benchmark	Count	1386	1580
	%	72%	67.2%
Total % at Benchmark	Count	1925	2351
	%	100%	100%

Findings: Second Grade

There were 5,620 matched DIBELS assessment scores for RF second grade students during the fall and spring testing periods.

DIBELS Measures
At the beginning and end of the year, second grade students were assessed on the ORF. The ORF fluency with connected text measure is the most important predictor of future student success in second grade.

Overall, the results show that the RF schools showed a slight gain in the percentage of second-grade students at benchmark on the DIBELS. On the ORF measure the students' mean beginning score was 33.72 compared to the average ORF end measure of 71.62; both the beginning and end mean scores are in the "some risk" category according to the DIBELS Benchmark performance criteria (Table 4.20). Approximately 45 percent of RF students were in the "at risk" categories at the beginning of the year, with 48 percent "at risk" at the end of the year. Conversely, 29 percent started the year at "low risk" and 32 percent achieved "low risk" by the end of year. Both the "at risk" and "low risk" categories gained in the number of students (Table 4.21).

Table 4.20
DIBELS Benchmark Performance Criteria

Second Grade	Category	Beginning of Year	End of Year
ORAL READING FLUENCY	At Risk	0 – 25	0 - 69
	Some Risk	26 – 43	70 – 89
	Low Risk	44+	90+

Table 4.21
Reading First - Second Grade Category Measures

READING FIRST	% Beginning of Year ORF	% End of Year ORF
Mean Score	33.72	71.62
At Risk	44.9 N = 2520	47.9 N = 2692
Some Risk	26.4 N = 1485	19.9 N = 1119
Low Risk	28.7 N = 1612	32.2 N = 1809
Total	100.0 N = 5617	100.0 N = 5620

In contrast, the comparison group lost ground with fewer students in the "low risk" category at the end of the year than the beginning (28.8% end compared to 37.5% beginning) – and with more students in the "at risk" category at the end than at the beginning (52.2% end compared to 37.9% beginning). As to the mean scores, the RF students caught up with and in fact slightly surpassed their counterparts in the pretest-posttest comparison schools.

Although the comparison group started slightly higher than the RF group, the latter surpassed the former by the end of second grade. In fact, 32.2 percent of RF students were at benchmark, compared to 28.8 and 24.9 percent for the other two groups at the end of the year. While there was a positive gain in the percent of RF students at "benchmark" at the end of year compared to the beginning of the year, the percentage of students at "benchmark" went in a negative direction for the comparison group (Table 4.22).

Table 4.22
Comparison Schools Category Measures

COMPARISON SCHOOLS	% Beginning of Year ORF	% End of Year ORF
Mean Score	39.38	68.05
At Risk	37.9 N = 331	52.2 N = 456
Some Risk	24.6 N = 215	19.0 N = 166
Low Risk	37.5 N = 327	28.8 N = 251
Total	100.0 N = 873	100.0 N = 873

Table 4.23
Post Only Schools Category Measures

POST ONLY SCHOOLS	% End of Year ORF
Mean Score	60.92
At Risk	58.2 N = 124
Some Risk	16.9 N = 36
Low Risk	24.9 N = 53
Total	100.0 N = 213

There were 47.9 percent of RF students at "intensive" at the end of the year compared to 52.2 percent of the comparison group and 58.2 percent of the Post Only group. The RF group increased only 3 percent in "intensive" category, whereas the comparison group increased by 14.3 percent in the intensive category; this is a move in the "wrong" direction. (Table 4.23)

Figure 4.6 shows the mean score for each of the categories for the second graders on the ORF measure. Therefore, for the "benchmark" students, the mean score at the beginning of year was 68.3 and at the end of the year it was 112.9, a growth of 44.6 points. Notice that the benchmark growth rate is a difference of 46 points (90 minus 44) meaning that the RF students improved just as much as might be anticipated.

Figure 4.6

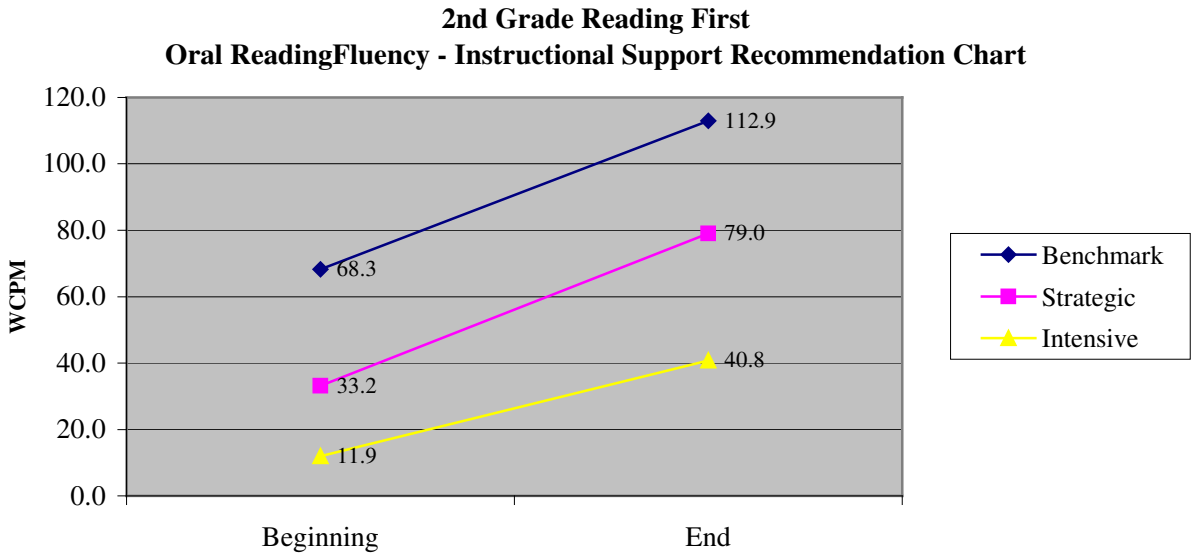
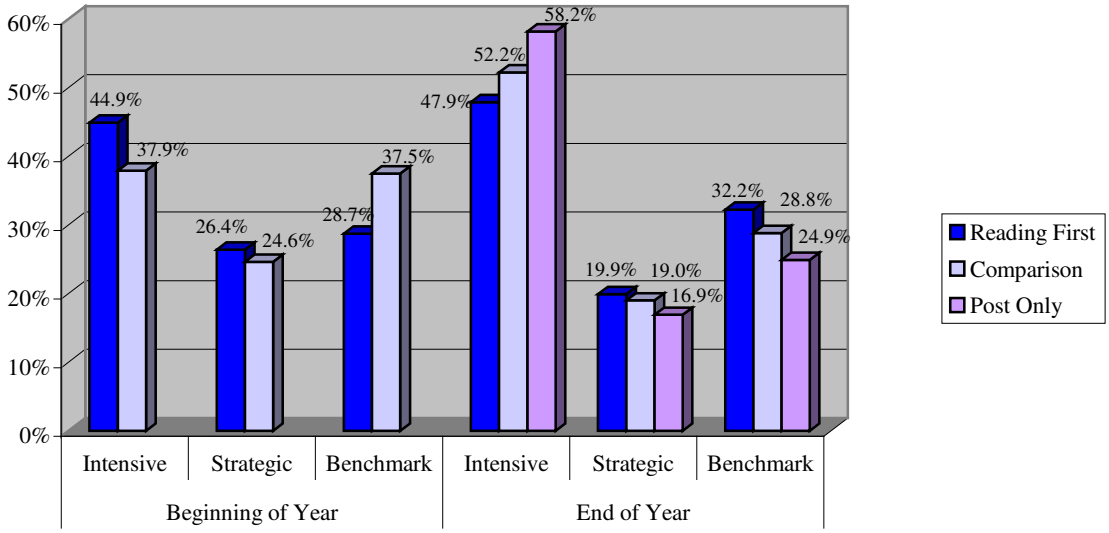


Table 4.24
Second Grade Instructional Support Recommendation

		Second Grade Instructional Recommendation – Beginning of Year				Second Grade Instructional Recommendation - End of Year			
		Intensive	Strategic	Benchmark	Total	Intensive	Strategic	Benchmark	Total
Reading First	N	2520	1485	1612	5617	2692	1119	1809	5620
	%	44.9%	26.4%	28.7%	100.0%	47.9%	19.9%	32.2%	100.0%
Comparison	N	331	215	327	873	456	166	251	873
	%	37.9%	24.6%	37.5%	100.0%	52.2%	19.0%	28.8%	100.0%
Post Only	N					124	36	53	213
	%					58.2%	16.9%	24.9%	100.0%
Total	N	2851	1700	1939	6490	3272	1321	2113	6706
	%	43.9%	26.2%	29.9%	100.0%	48.8%	19.7%	31.5%	100.0%

Figure 4.7

2nd Grade Instructional Support Recommendation



There was only a 3.4 percent difference between the RF and comparison groups of those who ended up at "benchmark" whereas there were 7.3 and 3.9 percent differences of RF and the comparison group to the end of year percent at "benchmark" for the Post only group (see Table 4.25).

Table 4.25
Second Grade Change Comparison End of Year Benchmark

Comparison End of Year Benchmark	% change
RF to Comparison group	3.4
RF to Post Only	7.3
Comparison to Post Only	3.9

Ethnicity

Table 4.26 shows the change in the benchmark category from the beginning to end of the year by ethnicity. According to the data, of the total 1,612 students at benchmark at the beginning of the year, 962 (59.7%) were Hispanic/Latino students and 347 (21.5%) were White, not Hispanic/Latino. By the end of the year, Hispanic/Latino students were 59.6 percent of the total benchmark students, while White, not Hispanic/Latino students were 21.2 percent of the total benchmark students.

Table 4.26
Beginning and End of the Year Instructional Support Recommendation
% at Benchmark by Ethnicity

		Second Grade - Beginning of Year ISR	Second Grade - End of Year ISR
		Benchmark	Benchmark
Not Reported	Count	137	165
	%	8.5%	9.1%
American Indian	Count	77	80
	%	4.8%	4.4%
Asian	Count	24	22
	%	1.5%	1.2%
Black/African-American, not Hispanic/Latino	Count	60	73
	%	3.7%	4.0%
Hispanic/Latino	Count	962	1079
	%	59.7%	59.6%
Native Hawaiian/Pacific Islander	Count	5	7
	%	0.3%	0.4%
Other	Count	0	0
	%	.0%	.0%
White, not Hispanic/Latino	Count	347	383
	%	21.5%	21.2%
Total	Count	1612	1809
	%	100%	100%

Free & Reduced Lunch
Table 4.27 shows the change in the benchmark category from the beginning to end of the year according to those students who qualify for free and reduced lunch. The data showed that of the total 1612 students at benchmark at the beginning of the year, 746 (46.3%) were eligible for free and reduced lunch. By the end of the year, students eligible for free and reduced lunch were 47.7 percent of the 1809 students at benchmark.

Table 4.27
Beginning and End of the Year Instructional Support Recommendation
% at Benchmark by Free & Reduced Lunch

		Second Grade - Beginning of Year ISR	Second Grade - End of Year ISR
		Benchmark	Benchmark
Not Reported	Count	635	704
	%	39.4%	38.9%
Eligible	Count	746	863
	%	46.3%	47.7%
Not Eligible	Count	231	242
	%	14.3%	13.4%
Total	Count	1612	1809
	%	100%	100%

Special Education
Table 4.28 shows the change in the benchmark category from the beginning to end of the year according to those students who qualify for special education. According to the data, of the total 1612 students at benchmark at the beginning of the year, 90 (5.6%) were eligible for special education. By the end of the year, students eligible for special education were 4.9 percent of the 1809 students at benchmark.

Table 4.28
Beginning and End of the Year Instructional Support Recommendation
% at Benchmark by Special Education

		Second Grade - Beginning of Year ISR	Second Grade - End of Year ISR
		Benchmark	Benchmark
Not Reported	Count	732	843
	%	45.4%	46.6%
Eligible	Count	90	88
	%	5.6%	4.9%
Not Eligible	Count	790	878
	%	49.0%	48.5%
Total	Count	1612	1809
	%	100%	100%

English Language Learners (ELL)

Table 4.29 shows the change in the benchmark category from the beginning to end of the year according to those students who qualify for ELL services. According to the data, of the total 1612 students at benchmark at the beginning of the year, 423 (26.2%) were ELL students. By the end of the year, ELL students were 27.4 percent of the 1809 students at benchmark.

Table 4.29
Beginning and End of the Year Instructional Support Recommendation
% at Benchmark by ELL

		Second Grade - Beginning of Year ISR	Second Grade - End of Year ISR
		Benchmark	**Benchmark**
ELL % at Benchmark	Count	423	496
	%	26.2%	27.4%
Non-ELL Students % at Benchmark	Count	1189	1313
	%	73.8%	72.6%
Total % at Benchmark	Count	1612	1809
	%	100%	100%

Findings: Third Grade

There were 5,328 matched DIBELS assessment scores for RF third grade students during the fall and spring testing periods.

DIBELS Measures
In the fall, third grade students were assessed on the RTF, WUF, and ORF (fluency with connected text). In the spring, students were assessed on the same measures. The fluency with connected text measure, in this case the ORF, is the most important predictor of future student success in this grade level.

Because both the category and ISR measures were based only on ORF scores, these data were the same. The DIBELS raw scores were coded into the "at risk," "some risk" and "low risk" categories. Approximately 46 percent or more of RF students were in the "at risk" categories at the beginning of the year, with 37 percent "at risk" at the end of the year. Conversely, 28 percent or less started the year at "low risk" with 31 percent achieving "low risk" by the end of year (Table 4.30). Both the RF and comparison group saw a downturn of "at risk" students. However, RF saw a more dramatic decline of 9.4 percent, while the decrease was a moderate 4.1 percent for the comparison group (Table 4.31). Even though the numbers for the Post Only were lower, they were within only 2 percent of both the RF and comparison group (Table 4.32).

Table 4.30
Reading First - Third Category Measures

READING FIRST	% Beginning of Year ORF	% End of Year ORF
Mean Score	57.81	89.60
At Risk	46.1 N = 2456	36.7 N = 1954
Some Risk	26.1 N = 1391	32.8 N = 1748
Low Risk	27.8 N = 1481	30.5 N = 1626
Total	100.0 N = 5328	100.0 N = 5328

Table 4.31
Comparison Schools Category Measures

COMPARISON SCHOOLS	% Beginning of Year ORF	% End of Year ORF
Mean Score	62.86	89.31
At Risk	41.6 N = 284	37.5 N = 256
Some Risk	26.4 N = 180	31.4 N = 214
Low Risk	32.0 N = 218	31.1 N = 212
Total	100.0 N = 682	100.0 N = 682

Table 4.32
Post Only Schools Category Measures

POST ONLY SCHOOLS	% End of Year ORF
Mean Score	88.71
At Risk	34.2 N = 67
Some Risk	36.7 N = 72
Low Risk	29.1 N = 57
Total	100.0 N = 196

DIBELS Instructional Support Recommendation

The Instructional Support Recommendation group also compared students. Again, the RF students started slightly lower, with 46.1 percent at the intensive category compared to 41.6 percent of the comparison group. By the end of the year, the RF students had caught up to the comparison group (36.7% vs. 37.5%). However, the RF students exceeded the end of year Post only group in benchmark by only 1.4 percent (Table 4.33).

Table 4.33
Third Grade Instructional Support Recommendation

		Third Grade Instructional Recommendation Beginning of Year				Third Grade Instructional Recommendation End of Year			
		Intensive	Strategic	Benchmark	Total	Intensive	Strategic	Benchmark	Total
Reading First	N	2456	1391	1481	5328	1954	1748	1626	5328
	%	46.1%	26.1%	27.8%	100.0%	36.7%	32.8%	30.5%	100.0%
Comparison	N	284	180	218	682	256	214	212	682
	%	41.6%	26.4%	32.0%	100.0%	37.5%	31.4%	31.1%	100.0%
Post Only	N					67	72	57	196
	%					34.2%	36.7%	29.1%	100.0%
Total	N	2740	1571	1699	6010	2277	2034	1895	6206
	%	45.6%	26.1%	28.3%	100.0%	36.7%	32.8%	30.5%	100.0%

Figure 4.9

3rd Grade Instructional Support Recommendation

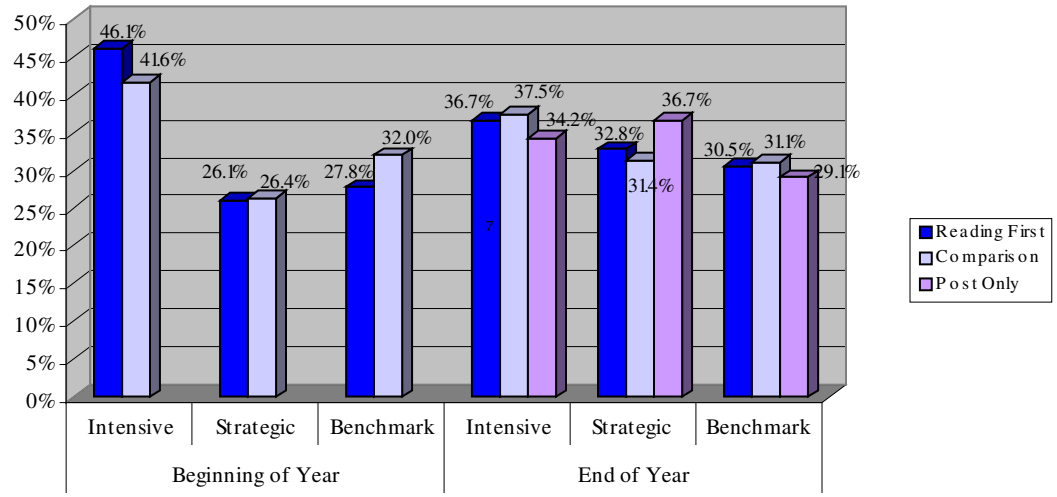
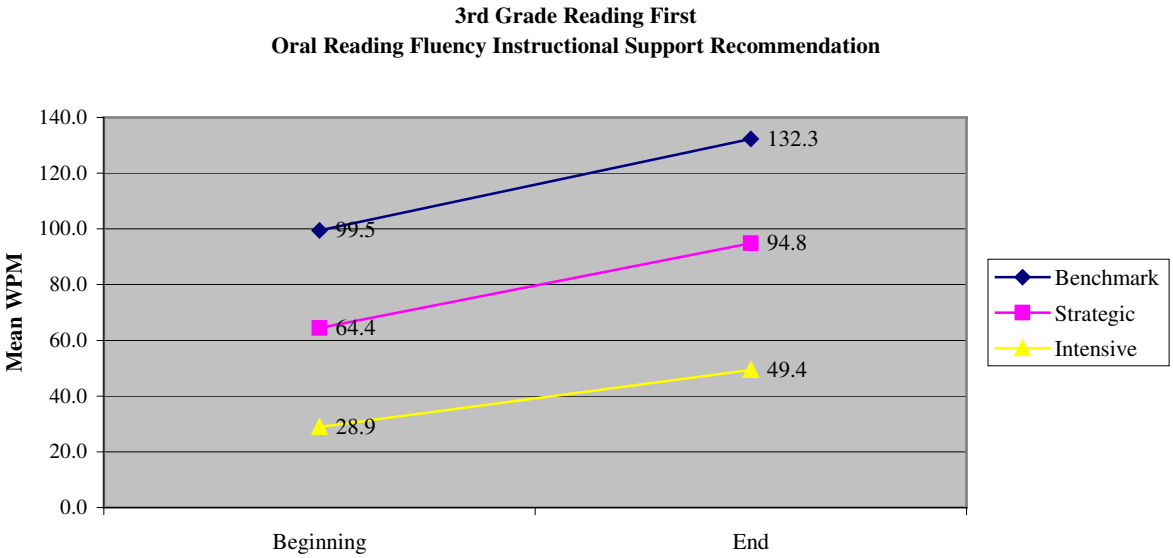


Figure 8 shows the mean score for the third grade Reading First students in each of the categories on the ORF measure. Therefore, for the "benchmark" students, the mean score at the beginning of year was 99.5 and at the end of the year it was 132.3, a growth of 32.8 points.

Figure 4.8



There was a –0.6 percent difference between the RF and comparison groups of those who ended up at "benchmark" whereas there were 1.4 and 2.0 percent differences of RF and the comparison group to the end of year percent at "benchmark" for the Post only group (see Table 4.34).

Table 4.34
Third Grade Change Comparison End of Year Benchmark

Comparison End of Year Benchmark	% change
RF to Comparison group	-0.6
RF to Post Only	1.4
Comparison to Post Only	2.0

Ethnicity

Table 4.35 shows the change in the benchmark category from the beginning to end of the year by ethnicity. According to the data, of the total 1,481 students at benchmark at the beginning of the year, 910 (61.4%) were Hispanic/Latino students and 296 (20%) were White, not Hispanic/Latino. By the end of the year, Hispanic/Latino students were 63 percent of the 1626 students at benchmark, while White, not Hispanic/Latino students were 18.5 percent of the 1626 students at benchmark.

Table 4.35
Beginning and End of the Year Instructional Support Recommendation
% at Benchmark by Ethnicity

		Third Grade - Beginning of Year ISR	Third Grade - End of Year ISR
		Benchmark	Benchmark
Not Reported	Count	125	155
	%	8.4%	9.5%
Alaska Native	Count	0	0
	%	.0%	.0%
American Indian	Count	75	72
	%	5.1%	4.4%
Asian	Count	26	26
	%	1.8%	1.6%
Black/African-American, not Hispanic/Latino	Count	36	36
	%	2.4%	2.2%
Hispanic/Latino	Count	910	1024
	%	61.4%	63.0%
Native Hawaiian/Pacific Islander	Count	13	12
	%	0.9%	0.7%
White, not Hispanic/Latino	Count	296	301
	%	20.0%	18.5%
Total	Count	1481	1626
	%	100%	100%

Free & Reduced Lunch
Table 4.36 shows the change in the benchmark category from the beginning to end of the year according to those students who qualify for free and reduced lunch. The data showed that of the total 1481 students at benchmark at the beginning of the year, 650 (43.9%) were eligible for free and reduced lunch. By the end of the year, students eligible for free and reduced lunch were 46.6 percent of the 1626 students at benchmark.

Table 4.36
Beginning and End of the Year Instructional Support Recommendation
% at Benchmark by Free & Reduced Lunch

		Third Grade - Beginning of Year ISR	Third Grade - End of Year ISR
		Benchmark	Benchmark
Not Reported	Count	637	664
	%	43.0%	40.8%
Eligible	Count	650	758
	%	43.9%	46.6%
Not Eligible	Count	194	204
	%	13.1%	12.5%
Total	Count	1481	1626
	%	100%	100%

Special Education
Table 4.37shows the change in the benchmark category from the beginning to end of the year according to those students who qualify for special education. According to the data, of the total 1481 students at benchmark at the beginning of the year, 76 (5.1%) were eligible for special education. By the end of the year, students eligible for special education were 5 percent of the 1626 students at benchmark.

Table 4.37
Beginning and End of the Year Instructional Support Recommendation
% at Benchmark by Special Education

		Third Grade - Beginning of Year ISR	Third Grade - End of Year ISR
		Benchmark	Benchmark
Not Reported	Count	739	813
	%	49.9%	50.0%
Eligible	Count	76	82
	%	5.1%	5.0%
Not Eligible	Count	666	731
	%	45.0%	45%
Total	Count	1481	1626
	%	100%	100%

English Language Learners (ELL)

Table 4.38 shows the change in the benchmark category from the beginning to end of the year according to those students who qualify for ELL services. According to the data, of the total 1481 students at benchmark at the beginning of the year, 325 (21.9%) were ELL students. By the end of the year, ELL students were 25.1 percent of the 1626 students at benchmark.

Table 4.38
Beginning and End of the year Instructional Support Recommendation
% at Benchmark by ELL

		Third Grade - Beginning of Year ISR	Third Grade - End of Year ISR
		Benchmark	**Benchmark**
ELL % at Benchmark	Count	325	408
	%	21.9%	25.1%
Non-ELL Students % at Benchmark	Count	1156	1218
	%	78.1%	74.9%
Total % at Benchmark	Count	1481	1626
	%	100%	100%

Movement into the "Benchmark" Grouping

Using the DIBELS, it was possible to track the movement of students out of the "intensive" or "strategic" groups into the "benchmark" group. It was also possible to determine whether students who began the year off at benchmark made enough gains over the year to remain at benchmark at the end of the year (spring). Table 4.40 summarizes information for each grade level on the students at benchmark at the end of the year and those students who moved into the benchmark category from the intensive and strategic categories.

Table 4.39 describes how to read the data in Table 4.40. For each grade level, there are two rows of information. The three cells in the top row report the percentage of students at that grade level who, according to their DIBELS scores, needed "intensive," "strategic" or "benchmark" instruction in reading at the end of the school year. There are also three cells in the lower row. The left-hand cell reports the percentage of students who were in the "intensive" group at the beginning of the year and moved up to "benchmark" by the end of the year. The middle cell reports the percentage who were in the "strategic" group at the beginning of the year and moved up to "benchmark" by the end of the year. The last cell reports the percentage of students who were in the "benchmark" group at the beginning of the year and remained there at the end.

Table 4.39
How to Read/Interpret the Change Benchmark Table

<table>
<tr><td>Percentage of Students Needing "Intensive" Intervention, according to the End of Year 04 DIBELS</td><td>Percentage of Students Needing "Strategic" Intervention, according to the End of Year 04 DIBELS</td><td>Percentage of Students at Benchmark, according to the End of Year 04 DIBELS</td><td rowspan="2">The difference between the percentage of students at Benchmark in the beginning of year and end of year. A positive score means that more students are at Benchmark at the end of year than were in the beginning. A negative score means the percentage of students at Benchmark declined.</td></tr>
<tr><td>Percentage of those Students Who Needed "Intensive" Intervention according to the Beginning '03 DIBELS who then hit Benchmark on the End '04 DIBELS
This is a measure of the effectiveness of the most intensive interventions in getting the lowest performing students to Benchmark this year.</td><td>Percentage of those Students Who Needed "Strategic" Intervention according to the Beginning '03 DIBELS who then hit Benchmark on the End '04 DIBELS
This is a measure of the effectiveness of the double-dosing/ interventions in getting the students who were somewhat below level in the beginning up to Benchmark this year.</td><td>Percentage of Students at Benchmark at Beginning '03 and still at Benchmark at End '04
This is a measure of the ability of your core program to keep students who started the at grade level progressing at level. Ideally, this figure would be 100%.</td></tr>
</table>

The larger cell in the far right column reports on the difference between the percentage of students in the "benchmark" group at the end of the year, compared to the beginning. A score of "+10," for example, would mean that an additional 10 percent of that grade's students were in the "benchmark" group at the end of the year. A score of "-10" would indicate that ten percent fewer students were in the "benchmark" group at the end of the year, compared to the beginning. Table 4.40 provides this information for all four primary grades.

Table 4.40
DIBELS Benchmark Results, 2003-04

Grade	N	% End Intensive % Beginning Intensive at Benchmark	% End Strategic % Beginning Strategic at Benchmark	% End Benchmark % Beginning Benchmark at Benchmark	% at Benchmark, Change from Beginning 03 to End 04
Kindergarten	5576	25.6	21.3	53.1	+44.0
All AZ RF Schools		41.0	65.1	89.5	
First Grade	5744	29.7	29.4	40.9	+7.4
All AZ RF Schools		14.0	39.0	72.8	
Second Grade	5617	47.9	19.9	32.2	+3.5
All AZ RF Schools		3.3	25.6	83.4	
Third Grade	5328	36.7	32.8	30.5	-+2.7
All AZ RF Schools		2.2	25.6	82.0	

Only in kindergarten was there strong success, across Reading First, in getting students out of the lower two groups and into the "benchmark" group. Fully 41 percent of those who were in the "intensive" group and 65 percent of those who started off in the "strategic" group ended up at benchmark by the end of the year. It was very positive that the overall percent saw an increase of 44 percent difference in benchmark, however still only slightly more than half of the kindergarten students reached benchmark by the end of the year (the goal is 100 percent).

In contrast, only 14 percent of first-grade students who started in the "intensive" group and about 39 percent of those in the "strategic" group moved up to benchmark by the end of the year. At both the second- and third-grade level, movement up to benchmark was even less common. Just 2 to 3 percent of those starting the year in the "intensive" group and 25 percent of those in the "strategic" group made sufficient gains to hit benchmark by spring.

The stronger results at kindergarten can probably be explained by two factors: first, it is easier, comparatively, to meet kindergarten benchmark standards than at other grades, and second, students in second and third grade have had more opportunity to get further behind and have more ground to make up in the same amount of time.

A kindergarten student beginning the year in the "intensive" group is unable, within the one-minute timing, to identify the initial sounds in more than four words (ISF) and/or can provide the name of fewer than two letters (LNF). To be in the benchmark group by the end of the year, the student must be able to provide the name of 40 letters (or 29 if all other tasks are successfully accomplished), identify 35 or more phonemes in words (fewer if all other tasks are successfully

accomplished) and decode 25 or more simple consonant-vowel-consonant nonsense words (again, fewer if the other two tasks are successfully mastered). This marks real growth in a student's phonemic awareness and phonics knowledge in spring, compared to the previous fall.

Yet the growth required for a first-grade student to move from "intensive" to "benchmark" is greater. A first-grade student in the "intensive" group in the fall can identify fewer than 25 letters (perhaps none), can separate out fewer than 10 phonemes from a series of short words (again, perhaps none), and can decode 12 or fewer simple nonsense words. To be at "benchmark" in the spring, this student must be able to identify 35 or more phonemes, decode 50 or more nonsense words, and read a grade-level text at a rate of 40 or more words correct per minute (again, the standard for one of these three tasks might be a little lower if the other two were successfully mastered). The growth required to make this transition is greater, particularly for students who truly knew no letter sounds and could not identify individual phonemes.

By second and third grade, students in the "intensive" group may need to make even greater progress to reach benchmark. Some of the students may be at the early first grade level, with little or emergent phonemic awareness and marginal decoding skills. In second grade, students in the "intensive" group start off reading 25 or fewer words correct per minute (wcpm); third-grade students read 52 or fewer wcpm. To hit the end-of-year benchmark targets, second-graders need to gain a minimum of 65 wcpm, while third-graders' rate needs to increase by 58 wcpm. These gains are nearly twice the "normal" gains made by students at grade level who make typical growth over the year.

Thus movement from "intensive" to "benchmark" grows increasingly challenging for the older grades. The only way to meet such challenges is to provide students with very systematic, individually targeted, and highly effective intervention.

Chapter 10 discusses the implementation of instructional changes during the first year of Reading First and notes that many schools struggled to put together and sustain a comprehensive intervention system. While some schools were able to identify students systematically and provide appropriate and highly targeted interventions, many schools were still asking questions in the spring about materials and delivery formats. It is very likely that the results at first, but especially at second and third grades, can be explained by the absence of a functioning intervention system across all schools. Without intensive and targeted interventions, students who were substantially behind in reading were not able to make enough gains in one year to catch up to their grade level.

Table XY also makes clear that, despite the adoption of core reading programs designed to provide systematic and explicit reading instruction, not all students who started the year at benchmark were still at benchmark at the end of the year. Instead, this rate varied from about 73 to 90 percent of them were (89.5 percent in kindergarten, 72.8 percent in first grade, 83.4 percent in second grade, and 82.0 percent in third grade).

At all grade levels, analyses were conducted to determine if there was a difference in the percentage of students meeting benchmark between the RF and eight comparison schools. The data show that at all points, RF schools had a slightly higher percentage of students at benchmark

at the end of the year than the comparison schools. Of the RF kindergarten students who started at "intensive," 41 percent moved to "benchmark" compared to only 26.9 percent of the comparison group. In first grade, 39 percent of "strategic" and 73 percent of "benchmark" moved to "benchmark" compared to only 15.7 and 47 percent respectively of the comparison group. The threnḍ was similar for both second and third grades, where a much larger percent of RF "strategic" and "benchmark" students moved to "benchmark" than was the case for the comparison second and third graders.

Table 4.41
DIBELS Comparison Pre-Post Benchmark Results, 2003-04

Grade	N	% End Intensive % Beginning Intensive at Benchmark	% End Strategic % Beginning Strategic at Benchmark	% End Benchmark % Beginning Benchmark at Benchmark	% at Benchmark, Change from Beginning 03 to End 04
Kindergarten	880	29.2	20.8	50.0	+30.9
Pre-Post Comparison Schools		26.9	58.3	86.9	
First Grade	938	40.2	28.9	30.9	-20.8
Pre-Post Comparison Schools		11.3	15.7	47.0	
Second Grade	873	52.2	19.0	28.8	-8.7
Pre-Post Comparison Schools		0.9	11.6	68.2	
Third Grade	682	37.5	31.4	31.1	-0.9
Pre-Post Comparison Schools		1.8	16.7	81.2	

Analysis of Covariance

The statistical method of choice in analyzing the data generated by the pretest and posttest control group design is analysis of covariance, in which the posttest measures are compared using pretest means as the covariate. For first, second and third grade data, one-way analysis of covariance was used.

It should be noted that due to the large number of cases and the extremely small effect size, the results are substantively not very informative. For the most part, this is true for all the grade levels. The quarterly report showed the analysis for the pretest and posttest comparison data using one-way analysis of variance of the pre-post "change scores."
This strategy of analysis was presented in the quarterly report for reasons noted; however, since the statistical method of choice is analysis of covariance, those results were presented herein as additional supportive data to the earlier analysis.

First Grade PSF
The ANCOVA was significant F = 23.39 (1 df, N = 6,684), p = .000, partial eta^2 =.003. However, the strength of the relationship between the group factor and the reading score was extremely small, as indicated by the partial eta squared or effect size, i.e., the group factor accounted for only .3% of the dependent variable.

The estimated marginal means, i.e., the means of groups adjusted for initial differences, indicated that the RF group had the largest adjusted mean (M = 49.05) as compared to the PrePost group (M = 46.84). The test of the covariate evaluates the relationship between the covariate and the dependent variable, controlling for the group factor F = 868.796, p = .000. In this case, the covariate accounted for 12% of the variance.

Second Grade ORF
The ANCOVA was significant F = 209.29 (1 df, N = 6,490), p = .000, partial eta^2 =.031. However, the strength of the relationship between the group factor and the reading score was extremely small, as indicated by the partial eta squared or effect size, i.e., the group factor accounted for 3% of the dependent variable.

The estimated marginal means, i.e., the means of groups adjusted for initial differences, indicated that the RF group had the largest adjusted mean (M = 72.51) as compared to the PrePost group (M = 62.51). The test of the covariate evaluates the relationship between the covariate and the dependent variable, controlling for the group factor F = 17395.46, p = .000. In this case, the covariate accounted for 73% of the variance.

Third Grade ORF
The ANCOVA was significant F = 57.173 (1 df, N = 6,010), p = .000, partial eta^2 =.009. However, the strength of the relationship between the group factor and the reading score was extremely small, as indicated by the partial eta squared or effect size, i.e., the group factor accounted for .9% of the dependent variable.

The estimated marginal means, i.e., the means of groups adjusted for initial differences, indicated that the RF group had the largest adjusted mean (M = 90.19) as compared to the PrePost group (M = 84.72). The test of the covariate evaluates the relationship between the covariate and the dependent variable, controlling for the group factor F = 22178, p = .000. In this case, the covariate accounted for 79% of the variance.

CHAPTER 5
STUDENT ASSESSMENTS II:
A CLOSER LOOK AT DIBELS SCHOOL-LEVEL VARIATION

Highlights

- Overall performance on the DIBELS assessment, as measured by the percentage of students at benchmark, varied considerably by school. It is crucial that the review of project-wide trends noted in the previous chapter be supplemented by an awareness of this school-level variation.

- It is possible for schools to show "negative growth" in the percentage of students at benchmark. This does not mean that student reading performance declined over the year, but rather that student improvement over the year was not enough for students to reach new, higher benchmarks in the spring. In kindergarten, no schools saw negative growth in the percentage of students at benchmark. At the other three grades, however, some schools did see this sort of decline.

- The evaluation team also conducted additional analyses of first-grade assessment items, looking specifically at the way in which student performance on the Nonsense Word Fluency (NWF, a test of decoding skill) at the beginning of the year predicted performance on the Oral Reading Fluency (ORF) at the end of the year. In general, performance on the NWF was highly predictive of the other; students who were "at risk" according to the NWF at the beginning usually had ORF scores substantially below (on average 15 wcpm lower) those who were in the "some risk" group.

- At 36 schools, the group of first-grade students "at risk" according to the beginning NWF had an average spring fluency score that was less than 15 points below the performance of the "some risk" group. This may suggest some success working to bring up the lowest students, although schools should individually analyze the movement of students in both their "at risk" and "some risk" group before drawing any firm conclusions.

School-level Variation in Attainment of Benchmark Status

The previous chapter presented overall trends in student performance, including the overall percentage of students in the "intensive", "strategic" and "benchmark" groups as well as the movement of students into the benchmark group. This chapter briefly examines the same information but presents school-level breakdowns and highlights the level of variation seen from school to school.

The previous chapter also explained how to read the data summary tables, but that explanation is repeated here and presented in Table 5.1. For each grade level, there are two rows of information. The three cells in the top row report the percentage of students at that grade level who, according to their DIBELS scores, needed "intensive," "strategic" or "benchmark" instruction in reading at the end of the school year. There are also three cells in the lower row.

The left-hand cell reports the percentage of students who were in the "intensive" group at the beginning of the year and moved up to "benchmark" by the end of the year. The middle cell reports the percentage who were in the "strategic" group at the beginning of the year and moved up to "benchmark" by the end of the year. The last cell reports the percentage of students who were in the "benchmark" group at the beginning of the year and remained there at the end.

The larger cell in the far right column reports on the difference between the percentage of students in the "benchmark" group at the end of the year, compared to the beginning. A score of "+10," for example, would mean that an additional 10 percent of that grade's students were in the "benchmark" group at the end of the year. A score of "-10" would indicate that ten percent fewer students were in the "benchmark" group at the end of the year, compared to the beginning.

Table 5.1
How to Read/Interpret the Change Benchmark Table

<table>
<tr><td>Percentage of Students Needing "Intensive" Intervention, according to the End of Year 04 DIBELS</td><td>Percentage of Students Needing "Strategic" Intervention, according to the End of Year 04 DIBELS</td><td>Percentage of Students at Benchmark, according to the End of Year 04 DIBELS</td><td rowspan="2">The difference between the percentage of students at Benchmark in the beginning of year and end of year (end or year minus beginning of year). A positive score means that more students are at Benchmark at the end of year than were in the beginning. A negative score means the percentage of students at Benchmark declined.</td></tr>
<tr><td>Percentage of those Students Who Needed "Intensive" Intervention according to the Beginning '03 DIBELS who then hit Benchmark on the End '04 DIBELS</td><td>Percentage of those Students Who Needed "Strategic" Intervention according to the Beginning '03 DIBELS who then hit Benchmark on the End '04 DIBELS</td><td>Percentage of Students at Benchmark at Beginning '03 and still at Benchmark at End '04</td></tr>
</table>

In the following section, separate tables and analysis are presented for each grade level.

Kindergarten

As reported in Chapter 4, the percentage of kindergarten students at benchmark increased 44 percent, from nine to 53 percent, between fall and spring. This overall finding can obscure the large differences among schools, however. While a few schools saw increases of over 70 percent, others saw little or no change in the percentage of students at benchmark. Thirty-two of the schools experienced an above-average increase in the percentage at benchmark end of year benchmark above 44 percent, while thirty-one had below-average changes. None of the schools had a decline in the percentage of kindergarteners at benchmark.

At the end of the year, slightly more than half of all Arizona Reading First kindergarten students (53.1%) had reached benchmark. Again, this statistic varied considerably by school, with a range from 22.9 at benchmark on the low end to 90.3 percent at the high end.

Table 5.2
Summary of Kindergarten Performance on the DIBELS, by School

	N	% End Intensive % Beginning Intensive at Benchmark	% End Strategic % Beginning Strategic at Benchmark	% End Benchmark % Beginning Benchmark at Benchmark	% at Benchmark, Change from Beginning 03 to End 04
All AZ RF Schools	5576	25.6 41.0	21.3 65.1	53.1 89.5	44.0
School, by District					
Alhambra	**610**				
Andalucia Primary School	204	23.0 30.3	31.9 60.6	45.1 92.9	38.2
Sevilla Primary School	186	31.7 31.2	22.0 63.3	46.2 82.4	37.1
Westwood School	220	20.5 41.4	28.6 63.5	50.9 100.0	45.4
Casa Grande	171				
Ironwood Elementary	71	9.9 76.7	5.6 95.2	84.5 100.0	74.6
Mesquite Elementary	100	28.0 36.4	16.0 75.0	56.0 92.3	43.0
Coolidge	175				
West Elementary School	175	26.3 43.1	18.9 62.5	54.9 100.0	45.2
Crane Schools	388				
H.L. Suverkrup	90	10.0 56.8	24.4 62.1	65.6 94.1	46.7
Pueblo	82	20.7 43.2	14.6 86.7	64.6 64.6	54.8
Salida Del Sol	107	53.3 22.0	21.5 27.3	25.2 100.0	22.4
Valley Horizon	109	13.8 38.6	23.9 68.2	62.4 100.0	43.1

Glendale Elementary	300				
Glendale American Elementary	88	19.3 40.7	21.6 59.0	59.1 81.8	34.1
Wm. C. Jack	212	42.0 24.4	27.4 48.9	30.7 57.1	27.4
Isaac	334				
J. B. Sutton	102	34.3 36.0	19.6 72.0	46.1 100.0	44.1
Mitchell	100	19.0 52.1	28.0 56.0	53.0 50.0	49.0
P. T. Coe	132	15.9 60.0	15.9 80.0	68.2 100.0	62.9
Liberty Elementary	58				
Rainbow Valley	58	13.8 50.0	12.1 85.2	74.1 90.9	55.1
Maricopa County	35				
Thomas J. Pappas Elementary	35	54.3 21.1	22.9 25.0	22.9 0.0	22.9
Mesa	536				
Hawthorne Elementary School	103	32.0 25.4	20.4 79.4	47.6 100.0	41.8
Holmes Elementary School	93	31.2 20.3	44.1 30.0	24.7 50.0	20.4
Lowell Elementary School	106	25.5 52.5	15.1 76.2	59.4 100.0	54.7
Roosevelt Elementary School	81	56.8 10.3	18.5 35.3	24.7 50.0	14.8
Whitman Elementary School	79	53.2 6.5	22.8 40.7	24.1 83.3	16.5
Whittier Elementary School	74	28.4 30.8	16.2 73.9	55.4 100.0	39.2
Nogales	196				
A.J. Mitchell	76	19.7 34.2	23.7 75.0	56.6 100.0	48.7
Challenger	76	11.8 60.0	17.1 84.6	71.1 100.0	64.5
Robert Bracker	44	27.3 37.5	20.5 90.0	52.3 100.0	47.8
Page	75				
Desert View Elementary	75	18.7 33.3	36.0 41.7	45.3 83.3	29.3
Parker	31				
LePera Elementary	31	 81.3	9.7 100.0	90.3 100.0	0.0
Pendergast Elementary	237				
Pendergast Elementary School	69	30.4 22.2	30.4 48.0	39.1 87.5	27.5
Westwind Primary	168	33.3 38.4	18.5 63.0	48.2 90.0	42.2

Red Mesa	57				
Red Mesa Elementary School	55	10.9	23.6	65.5	41.9
		64.3	57.1	76.9	
Round Rock Elementary School	2		50.0	50.0	0.0
			.0	100.0	
Roosevelt	384				
C. J. Jorgensen School	72	23.6	13.9	62.5	56.9
		47.7	83.3	100.0	
J. R. Davis School	57	8.8	19.3	71.9	66.6
		68.8	77.3	66.7	
M. O. Bush School	74	6.8	6.8	86.5	78.4
		76.7	100.0	100.0	
Southwest School	70	10.0	14.3	75.7	61.4
		76.9	61.9	100.0	
Sunland School	62	25.8	29.0	45.2	45.2
		43.9	47.6	0.0	
T. G. Barr School	49	16.3	20.4	63.3	51.1
		36.4	85.7	83.3	
Safford	67				
Lafe Nelson School	67	31.3	25.4	43.3	28.4
		11.5	54.8	90.0	
Somerton	212				
Desert Sonora	59	47.5	22.0	30.5	28.8
		20.0	50.0	100.0	
Orange Grove	46	15.2	13.0	71.7	71.7
		69.7	76.9	0.0	
Tierra del Sol	107	44.9	29.9	25.2	23.3
		17.1	41.4	100.0	
Stanfield	82				
Stanfield School District 24	82	22.0	22.0	56.1	48.8
		50.8	66.7	83.3	
Sunnyside	253				
Craycroft Elementary	78	16.7	10.3	73.1	59.0
		65.5	83.3	100.0	
Drexel Elementary	92	12.0	18.5	69.6	67.4
		65.8	82.4	100.0	
Summit View Elementary	83	20.5	26.5	53.0	47.0
		38.5	73.1	100.0	
Tempe	257				
Curry School	103	14.6	14.6	70.9	43.7
		51.4	73.7	92.9	
Evans School	65	9.2	3.1	87.7	73.9
		80.6	96.0	88.9	
Laird School	89	22.5	19.1	58.4	44.9
		43.1	73.1	91.7	

Tolleson	70				
P.H. Gonzales	70	5.7	20.0	74.3	68.6
		64.1	88.9	75.0	
Tucson	388				
C.E. Rose Elementary	73	20.5	17.8	61.6	53.4
		56.8	69.6	66.7	
Davidson Elementary	49	36.7	20.4	42.9	28.6
		31.3	38.5	85.7	
Lynn-Uriquides Elementary	113	11.5	17.7	70.8	65.5
		61.4	83.8	100.0	
Menlo Park Elementary	39	30.8	20.5	48.7	41.0
		41.4	57.1	100.0	
Pueblo Gardens Elementary	53	30.2	34.0	35.8	15.0
		26.1	31.6	63.6	
Roberts Elementary	61	29.5	9.8	60.7	54.1
		45.7	77.3	100.0	
Washington	248				
Mountain View	116	37.1	29.3	33.6	31.0
		24.1	57.7	100.0	
Shaw Butte	132	52.3	20.5	27.3	24.3
		14.8	42.6	100.0	
Wickenburg	65				
MacLennan Elementary	65	12.3	16.9	70.8	40.0
		50.0	82.6	80.0	
Willcox	94				
Willcox Elementary School	94	27.7	22.3	50.0	35.1
		29.2	62.5	92.9	
Yuma Elementary	251				
Gwyneth Ham Elementary School	46	10.9	10.9	78.3	67.4
		69.6	83.3	100.0	
Palmcroft Elementary School	114	19.3	18.4	62.3	50.9
		47.1	68.0	100.0	
Roosevelt Elementary School	91	38.0	23.1	46.2	37.4
		32.7	61.3	75.0	

First Grade

Chapter 4 reported that across all Arizona Reading First schools, there was an increase in the percentage of first-grade students at benchmark in the spring, compared to the previous fall. Once again this overall average obscured the considerable variation that existed by school. In fact, thirteen schools actually saw a decline in the percentage of students at benchmark. On the other end, a few schools saw positive gains of more than 30 percent more students at benchmark, and quite a number saw increases of over 20 percent. These results are presented in Table 5.3.

At the end of the year, about four out of ten of the Arizona Reading First first-grade students (40.9%) had reached benchmark. Thirty schools succeeded in getting more students than average to benchmark, in one case getting as many as 74.4 percent to benchmark. Thirty-three schools had fewer than 40.9 percent of their first-grade students at benchmark, with a minimum of 3.8 percent at benchmark.

Table 5.3
Summary of First Grade Performance on the DIBELS, by School

	N	% End Intensive % Beginning Intensive at Benchmark	% End Strategic % Beginning Strategic at Benchmark	% End Benchmark % Beginning Benchmark at Benchmark	% at Benchmark, Change from Beginning 03 to End 04
All AZ RF Schools	5744	29.7 14.0	29.4 39.0	40.9 72.8	7.4
School, by District					
Alhambra	624				
Andalucia Primary School	237	21.5 18.8	26.2 30.2	52.3 69.7	-7.6
Sevilla Primary School	195	43.6 0.0	27.7 21.2	28.7 59.2	-10.3
Westwood School	192	28.1 9.5	33.3 25.0	39.5 64.6	-4.2
Casa Grande	178				
Ironwood Elementary	81	13.6 41.5	25.9 75.0	60.5 87.5	40.7
Mesquite Elementary	97	16.5 26.0	35.1 59.3	48.5 90.0	27.9
Coolidge Unified	180				
West Elementary School	180	23.9 11.1	34.4 30.5	41.7 76.1	4.5
Crane Schools	464				
H.L. Suverkrup	97	19.6 20.0	36.1 33.3	44.3 62.5	-5.2
Pueblo	92	20.7 20.0	25.0 51.6	54.3 90.3	20.6
Salida Del Sol	145	46.2 9.5	29.7 32.0	24.1 57.1	9.6
Valley Horizon	130	16.9 16.0	23.8 54.8	59.2 75.7	2.3

Glendale Elementary	302				
Glendale American Elementary	97	38.1 0.0	11.3 38.9	50.5 76.4	-6.2
Wm. C. Jack	205	44.9 12.1	27.8 30.0	27.3 65.8	8.8
Isaac	358				
J. B. Sutton	122	38.5 14.3	25.4 46.2	36.1 60.0	3.3
Mitchell	108	20.4 11.5	29.6 34.8	50.0 72.9	-4.6
P. T. Coe	128	12.5 21.1	27.3 51.3	60.2 75.7	5.5
Liberty Elementary	48				
Rainbow Valley	48	33.3 5.9	31.3 37.5	35.4 66.7	4.1
Maricopa County	52				
Thomas J. Pappas Elementary	52	84.6 0.0	11.5 0.0	3.8 15.4	-21.2
Mesa	517				
Hawthorne Elementary School	90	27.8 30.2	22.2 77.8	50.0 78.9	28.9
Holmes Elementary School	91	12.1 13.3	31.9 43.8	56.0 68.6	7.6
Lowell Elementary School	131	42.0 7.0	26.0 27.3	32.1 70.7	0.8
Roosevelt Elementary School	58	12.1 20.0	31.0 53.8	56.9 76.7	5.2
Whitman Elementary School	81	27.2 6.9	29.6 37.5	43.2 85.7	8.6
Whittier Elementary School	66	31.8 5.6	22.7 42.1	45.5 72.4	1.6
Nogales	206				
A.J. Mitchell	88	6.8 38.9	34.1 52.2	59.1 89.7	26.1
Challenger	79	24.1 22.2	30.4 26.7	45.6 77.4	6.4
Robert Bracker	39	2.6 0.0	33.3 75.0	64.1 90.9	35.9
Page	85				
Desert View Elementary	85	27.1 21.4	32.9 25.8	40.0 76.9	9.4
Parker	39				
LePera Elementary	39	2.6 0.0	23.1 50.0	74.4 79.4	-12.8
Pendergast Elementary	259				
Pendergast Elementary School	87	48.3 4.5	28.7 22.2	23.0 56.0	-5.7
Westwind Primary	172	30.8 10.4	29.7 40.0	39.5 74.5	7.5

Red Mesa	48				
Red Mesa Elementary School	34	26.5 0.0	44.1 0.0	29.4 52.6	-26.5
Round Rock Elementary School	14	7.1	42.9 25.0	50.0 60.0	-21.4
Roosevelt	363				
C. J. Jorgensen School	67	31.3 17.1	31.3 47.6	37.3 81.8	20.9
J. R. Davis School	58	32.8 5.3	32.8 40.0	34.5 57.9	1.7
M. O. Bush School	59	23.7 8.3	16.9 64.3	59.3 75.8	3.4
Southwest School	76	25.0 19.2	21.1 52.6	53.9 83.9	13.1
Sunland School	57	36.8 0.0	35.1 54.5	28.1 47.6	-8.7
T. G. Barr School	46	37.0 10.5	28.3 53.3	34.8 50.0	8.7
Safford	71				
Lafe Nelson School	71	33.8 12.8	36.6 33.3	29.6 71.4	9.9
Somerton	218				
Desert Sonora	76	21.1 30.0	39.5 30.0	39.5 57.7	5.3
Orange Grove	46	23.9 25.0	34.8 42.1	41.3 63.6	17.4
Tierra del Sol	96	21.9 24.2	28.1 42.3	50.0 78.4	11.5
Stanfield	74				
Stanfield School District 24	74	32.4 15.6	35.1 28.6	32.4 78.6	13.5
Sunnyside	275				
Craycroft Elementary	89	24.7 24.0	37.1 45.5	38.2 70.6	19.1
Drexel Elementary	105	27.6 14.9	38.1 62.5	34.3 78.6	21.0
Summit View Elementary	81	30.9 14.3	30.9 47.6	38.3 83.3	16.1
Tempe	225				
Curry School	82	25.6 0.0	28.0 21.4	46.3 84.2	0.0
Evans School	67	34.3 4.5	28.4 19.0	37.3 83.3	1.5
Laird School	76	34.2 4.3	19.7 38.5	46.1 88.9	10.6
Tolleson	76				
P.H. Gonzales	76	40.8 9.1	31.6 42.9	27.6 72.7	13.1

Tucson	410				
C.E. Rose Elementary	86	25.6	37.2	37.2	30.2
		20.7	72.7	66.7	
Davidson Elementary	45	28.9	24.4	46.7	13.4
		20.0	50.0	80.0	
Lynn-Uriquides Elementary	99	17.2	31.3	51.5	20.2
		14.8	53.7	80.6	
Menlo Park Elementary	54	46.3	24.1	29.6	7.4
		4.0	17.6	100.0	
Pueblo Gardens Elementary	39	43.6	28.2	28.2	10.3
		11.5	33.3	85.7	
Roberts Elementary	87	32.2	31.0	36.8	27.6
		20.8	53.8	87.5	
Washington	281				
Mountain View	143	48.3	32.2	19.6	7.0
		6.2	22.7	72.2	
Shaw Butte	138	58.0	16.7	25.4	13.1
		7.7	50.0	76.5	
Wickenburg	71				
MacLennan Elementary	71	8.5	23.9	67.6	8.4
		0.0	57.1	85.7	
Willcox	85				
Willcox Elementary School	85	40.0	29.4	30.6	14.1
		13.2	27.3	85.7	
Yuma Elementary	235				
Gwyneth Ham Elementary School	56	26.8	39.3	33.9	14.3
		21.1	30.8	63.6	
Palmcroft Elementary School	98	16.3	39.8	43.9	-6.1
		19.2	34.8	61.2	
Roosevelt Elementary School	81	29.6	38.3	32.1	11.1
		19.4	25.0	70.6	

Second Grade

In the spring, 32.2 percentage of all Arizona Reading First second-grade students were at benchmark, an increase of 3.5 percentage points compared to the fall; this information was also presented in the previous chapter. Results were very different for the individual schools however. Some saw double-digit gains, while eighteen schools had fewer students at benchmark in the spring than in the fall.

While overall only about one-third of the second grade students (32.2%) had reached benchmark by the end of the year, four schools succeeded in getting over half of their students to benchmark. Thirty schools had fewer than 32.2 percent at benchmark, with the minimum being 15.6 percent. These results are summarized in Table 5.4.

Table 5.4
Summary of Second Grade Performance on the DIBELS, by School

	N	% End Intensive % Beginning Intensive at Benchmark	% End Strategic % Beginning Strategic at Benchmark	% End Benchmark % Beginning Benchmark at Benchmark	% at Benchmark, Change from Beginning 03 to End 04
All AZ RF Schools	5617	47.9 3.3	19.9 25.6	32.2 83.4	3.5
School, by District					
Alhambra	586				
Andalucia Primary School	209	56.9 1.2	20.6 12.1	22.5 64.4	-5.7
Sevilla Primary School	181	61.3 1.1	17.7 4.9	21.0 68.6	-7.2
Westwood School	196	54.1 3.4	16.8 17.9	29.1 84.6	2.6
Casa Grande	152				
Ironwood Elementary	56	44.6 12.5	14.3 70.0	41.1 85.7	16.1
Mesquite Elementary	96	29.2 10.5	16.7 72.4	54.2 93.1	24.0
Coolidge	151				
West Elementary School	151	31.1 5.6	26.5 27.9	42.4 83.3	6.6
Crane	457				
H.L. Suverkrup	93	32.3 0.0	23.7 23.1	44.1 85.4	0
Pueblo	116	24.1 9.5	19.0 25.0	56.9 88.9	2.6
Salida Del Sol	126	61.1 0.0	17.5 13.9	21.4 75.9	-1.6
Valley Horizon	122	32.0 8.0	18.9 6.5	49.2 84.8	-4.9

Glendale Elementary	319				
Glendale American Elementary	118	41.5 0.0	21.2 15.0	37.3 75.9	-8.5
Wm. C. Jack	201	59.2 0.9	21.4 20.8	19.4 77.8	1.5
Isaac	334				
J. B. Sutton	113	55.8 0.0	18.6 15.4	25.7 78.1	-2.6
Mitchell	97	51.5 0.0	15.5 3.3	33.0 83.8	-5.1
P. T. Coe	124	49.2 1.8	21.8 34.2	29.0 73.3	4.8
Liberty Elementary	39				
Rainbow Valley	39	46.2 0.0	17.9 20.0	35.9 92.3	2.6
Maricopa County	53				
Thomas J. Pappas Elementary	53	66.0 5.0	13.2 25.0	20.8 88.9	3.8
Mesa	541				
Hawthorne Elementary School	104	33.7 14.5	26.0 47.8	40.4 88.5	15.4
Holmes Elementary School	112	37.5 8.9	27.7 29.4	34.8 75.8	5.3
Lowell Elementary School	116	41.4 5.0	21.6 42.3	37.1 96.7	11.2
Roosevelt Elementary School	65	33.8 5.6	21.5 20.0	44.6 88.9	3.1
Whitman Elementary School	73	30.1 4.3	24.7 30.4	45.2 92.6	8.2
Whittier Elementary School	71	21.1 13.3	19.7 47.8	59.2 87.9	12.7
Nogales	216				
A.J. Mitchell	88	38.6 0.0	29.5 20.0	31.8 77.8	1.1
Challenger	82	35.4 0.0	19.5 39.3	45.1 92.9	11.0
Robert Bracker	46	45.7 0.0	26.1 38.9	28.3 85.7	13.1
Page	75				
Desert View Elementary	75	57.3 0.0	18.7 8.0	24.0 76.2	-4.0
Parker	36				
LePera Elementary	36	38.9 6.7	27.8 30.8	33.3 87.5	11.1
Pendergast	249				
Pendergast Elementary School	92	52.2 2.4	18.5 14.3	29.3 79.3	-2.2
Westwind Primary	157	50.3 1.4	17.2 15.4	32.5 89.8	1.3

Red Mesa	54				
Red Mesa Elementary School	39	64.1	17.9	17.9	-5.2
		0.0	5.0	66.7	
Round Rock Elementary School	15	60.0	6.7	33.3	0
		0.0	0.0	100.0	
Roosevelt	378				
C. J. Jorgensen School	76	63.2	13.2	23.7	7.9
		2.4	27.3	91.7	
J. R. Davis School	62	56.5	14.5	29.0	-1.6
		0.0	0.0	94.7	
M. O. Bush School	71	46.5	18.3	35.2	9.8
		2.6	57.1	88.9	
Southwest School	51	47.1	23.5	29.4	-3.9
		0.0	7.7	82.4	
Sunland School	72	62.5	13.9	23.6	1.4
		0.0	22.2	81.3	
T. G. Barr School	46	41.3	19.6	39.1	6.5
		11.8	14.3	93.3	
Safford	61				
Lafe Nelson School	61	27.9	23.0	49.2	9.9
		13.6	40.0	87.5	
Somerton	244				
Desert Sonora	60	50.0	23.3	26.7	-6.6
		4.2	0.0	75.0	
Orange Grove	64	70.3	14.1	15.6	3.1
		0.0	23.1	87.5	
Tierra del Sol	120	50.8	20.0	29.2	3.4
		0.0	32.1	83.9	
Stanfield	66				
Stanfield School District 24	66	60.6	21.2	18.2	12.1
		0.0	56.3	75.0	
Sunnyside	246				
Craycroft Elementary	74	48.6	17.6	33.8	9.5
		7.5	43.8	83.3	
Drexel Elementary	85	51.8	21.2	27.1	13.0
		10.0	46.2	91.7	
Summit View Elementary	87	48.3	18.4	33.3	17.2
		9.6	47.6	100.0	
Tempe	233				
Curry School	107	45.8	12.1	42.1	4.7
		2.5	25.9	92.5	
Evans School	67	62.7	13.4	23.9	10.5
		0.0	41.2	100.0	
Laird School	59	22.0	20.3	57.6	8.4
		11.8	30.8	96.6	
Tolleson	79				
P.H. Gonzales	79	41.8	25.3	32.9	3.8
		0.0	16.7	95.7	

Tucson	630				
C.E. Rose Elementary	68	47.1	19.1	33.8	17.6
		12.8	60.0	100.0	
Davidson Elementary	43	39.5	23.3	37.2	-2.3
		0.0	0.0	94.1	
Lynn-Uriquides Elementary	111	55.0	21.6	23.4	-4.5
		0.0	10.7	74.2	
Menlo Park Elementary	53	58.5	20.8	20.8	-3.7
		0.0	11.1	76.9	
Pueblo Gardens Elementary	48	47.9	18.8	33.3	18.7
		4.2	47.1	100.0	
Roberts Elementary	74	48.6	21.6	29.7	12.1
		6.5	53.3	84.6	
Washington	252				
Mountain View	127	65.4	17.3	17.3	4.7
		0.0	38.5	75.0	
Shaw Butte	125	64.8	14.4	20.8	0
		1.2	25.0	80.8	
Wickenburg	61				
MacLennan Elementary	61	32.8	29.5	37.7	-4.9
		0.0	19.0	73.1	
Willcox	90				
Willcox Elementary School	90	40.0	20.0	40.0	7.8
		3.0	28.6	93.1	
Yuma	248				
Gwyneth Ham Elementary School	69	52.2	13.0	34.8	8.7
		3.3	33.3	88.9	
Palmcroft Elementary School	74	44.6	25.7	29.7	-5.4
		0.0	13.0	73.1	
Roosevelt Elementary School	105	50.5	26.7	22.9	1.9
		2.0	29.4	59.1	

Third Grade

Between fall and spring, an additional 2.7 percent of Arizona Reading First third-grade students attained benchmark status on the DIBELS. Just as at the other grade levels, this overall average concealed dramatic variation by school.

In third grade, the overall percent difference in the benchmark group from the beginning to the end of the year saw a 2.7 percent positive difference. A few schools saw very strong gains of 18 percentage points or more, but seventeen schools had fewer students at benchmark in the spring than in the fall.

While overall only about three out of ten of the third-grade students (30.5%) reached benchmark, five schools managed to get over half of their students to benchmark. Thirty-four schools had fewer than 30.5 percent of their students at benchmark; the minimum was 11.8 percent.

Table 5.5
Summary of Third Grade Performance on the DIBELS, by School

	N	% End Intensive % Beginning Intensive at Benchmark	% End Strategic % Beginning Strategic at Benchmark	% End Benchmark % Beginning Benchmark at Benchmark	% at Benchmark, Change from Beginning 03 to End 04
All AZ RF Schools	5328	36.7 2.2	32.8 25.6	30.5 82.0	2.7
School, by District					
Alhambra	616				
Andalucia Primary School	213	31.9 0.0	33.8 11.4	34.3 59.5	-20.2
Sevilla Primary School	199	44.2 0.0	32.2 9.4	23.6 72.4	-5.5
Westwood School	204	32.4 0.0	37.7 29.2	29.9 91.3	7.4
Casa Grande	153				
Ironwood Elementary	71	39.4 2.4	31.0 38.5	29.6 88.2	5.7
Mesquite Elementary	82	24.4 0.0	23.2 52.6	52.4 100.0	12.2
Coolidge					
West Elementary School					NR
Crane	480				
H.L. Suverkrup	91	26.4 0.0	38.5 17.2	35.2 81.8	-1.1
Pueblo	108	17.6 9.1	28.7 35.1	53.7 87.8	8.3
Salida Del Sol	132	43.2 0.0	35.6 23.7	21.2 82.6	3.8
Valley Horizon	149	21.5 0.0	38.3 14.6	40.3 85.5	-1.3

Glendale Elementary	307				
Glendale American Elementary	120	30.8 0.0	34.2 8.1	35.0 75.0	-8.3
Wm. C. Jack	187	47.1 0.9	31.0 26.2	21.9 74.4	1.0
Isaac	324				
J. B. Sutton	115	44.3 3.3	30.4 25.0	25.2 76.9	2.6
Mitchell	95	37.9 2.1	29.5 6.7	32.6 90.6	-1.1
P. T. Coe	114	40.4 1.7	33.3 27.3	26.3 90.9	7.0
Liberty Elementary	56				
Rainbow Valley	56	23.2 0.0	41.1 26.7	35.7 100.0	7.1
Maricopa County	47				
Thomas J. Pappas Elementary	47	51.1 3.3	21.3 55.6	27.7 87.5	10.7
Mesa	476				
Hawthorne Elementary School	81	28.4 14.9	33.3 47.1	38.3 94.1	17.3
Holmes Elementary School	87	31.0 5.7	31.0 29.2	37.9 85.7	5.7
Lowell Elementary School	80	25.0 10.3	30.0 63.6	45.0 94.7	21.2
Roosevelt Elementary School	64	14.1 7.1	34.4 14.3	51.6 83.3	-4.7
Whitman Elementary School	92	14.1 10.3	40.2 14.3	45.7 85.7	0
Whittier Elementary School	72	25.0 3.8	22.2 55.6	52.8 96.4	13.9
Nogales	219				
A.J. Mitchell	95	27.4 0.0	42.1 41.9	30.5 94.1	12.6
Challenger	85	28.2 5.4	34.1 37.5	37.6 87.5	9.4
Robert Bracker	39	12.8 8.3	30.8 60.0	56.4 100.0	25.6
Page	71				
Desert View Elementary	71	56.3 0.0	26.8 23.5	16.9 72.7	1.4
Parker	37				
LePera Elementary	37	29.7 0.0	40.5 23.1	29.7 57.1	-8.1
Pendergast	239				
Pendergast Elementary School	84	35.7 2.9	27.4 16.7	36.9 87.1	0
Westwind Primary	155	42.6 0.0	34.8 17.0	22.6 75.0	-0.6

Red Mesa	52				
Red Mesa Elementary School	35	22.9 0.0	54.3 0.0	22.9 80.0	-5.7
Round Rock Elementary School	17	17.6 0.0	70.6 20.0	11.8 33.3	-5.8
Roosevelt	362				
C. J. Jorgensen School	76	50.0 0.0	25.0 30.4	25.0 92.3	7.9
J. R. Davis School	53	56.6 0.0	26.4 36.4	17.0 50.0	-1.9
M. O. Bush School	61	42.6 13.8	23.0 25.0	34.4 81.3	8.2
Southwest School	73	35.6 0.0	32.9 20.8	31.5 85.7	2.7
Sunland School	54	44.4 0.0	40.7 11.1	14.8 66.7	-1.9
T. G. Barr School	45	53.3 6.1	26.7 42.9	20.0 80.0	8.9
Safford	54				
Lafe Nelson School	54	33.3 0.0	38.9 9.1	27.8 70.0	-9.2
Somerton	250				
Desert Sonora	67	37.3 0.0	38.8 9.1	23.9 70.0	-6.0
Orange Grove	67	40.3 0.0	25.4 23.5	34.3 86.4	1.5
Tierra del Sol	116	39.7 1.7	41.4 27.8	19.0 55.0	1.8
Stanfield	70				
Stanfield School District 24	70	41.4 8.9	28.6 50.0	30.0 84.6	11.4
Sunnyside	215				
Craycroft Elementary	59	40.7 3.2	35.6 7.1	23.7 85.7	0
Drexel Elementary	82	51.2 0.0	34.1 11.1	14.6 83.3	0
Summit View Elementary	74	52.7 4.3	18.9 42.9	28.4 100.0	10.8
Tempe	205				
Curry School	81	35.8 3.7	27.2 9.1	37.0 84.4	-2.5
Evans School	52	28.8 5.0	36.5 25.0	34.6 81.3	3.8
Laird School	72	30.6 0.0	29.2 30.0	40.3 92.0	5.6
Tolleson	71				
P.H. Gonzales	71	35.2 5.4	32.4 26.7	32.4 89.5	5.6

Tucson	377				
C.E. Rose Elementary	66	40.9	25.8	33.3	18.1
		7.7	58.8	90.0	
Davidson Elementary	46	50.0	21.7	28.3	6.6
		0.0	37.5	100.0	
Lynn-Uriquides Elementary	115	34.8	33.9	31.3	4.3
		0.0	30.0	87.1	
Menlo Park Elementary	42	50.0	38.1	11.9	-7.1
		0.0	0.0	62.5	
Pueblo Gardens Elementary	41	41.5	34.1	24.4	9.8
		0.0	50.0	83.3	
Roberts Elementary	67	46.3	35.8	17.9	6.0
		0.0	62.5	87.5	
Washington	235				
Mountain View	128	41.4	32.0	26.6	4.7
		3.0	30.3	78.6	
Shaw Butte	107	54.2	27.1	18.7	0.9
		.0	17.2	78.9	
Wickenburg					
MacLennan Elementary					
Hassayampa Upper Elementary	87	21.8	34.5	43.7	0
		0.0	19.0	89.5	
Willcox	79				
Willcox Elementary School	79	48.1	30.4	21.5	1.2
		0.0	16.7	87.5	
Yuma Elementary	246				
Gwyneth Ham Elementary School	69	30.4	40.6	29.0	10.2
		3.0	30.4	92.3	
Palmcroft Elementary School	84	48.8	29.8	21.4	5.9
		0.0	20.6	84.6	
Roosevelt Elementary School	93	38.7	35.5	25.8	3.2
		0.0	19.0	95.2	

The Relationship Between NWF and ORF Scores in First Grade

In addition to the traditional analyses, evaluators also responded to a request from Arizona Reading First project staff to look more closely at the relationship between two measures on the first-grade DIBELS, the fall Nonsense Word Fluency (NWF) measure and the spring Oral Reading Fluency (ORF) measure. In particular, staff members were curious about the ability of fall NWF scores to predict spring ORF scores. In fact, Arizona student data did in fact fit the predicted linear model; the technical analysis for how this was determined is described in Appendix B. NWF and ORF mean scores by category by school were tabulated, analyzed and presented in the summer quarterly report. The school-level results are included in Appendix C of this report. This section simply summarizes some of the school findings more generally.

The DIBELS (NWF) is a quick assessment of decoding skills, while the ORF assesses fluency of reading unfamiliar text. The ORF is not administered until the spring of first grade simply because first grade is the year that most students develop enough reading skill to test fluency.

In the fall, 43.3 percent of Arizona Reading First students were "at risk" on the NWF. In the spring, 29.7 percent of students were "at risk" on the ORF.

In response to inquiries from the Arizona Reading First staff, an analysis was conducted to examine what happened to those students who started the year "at risk" on the NWF, 43.3 percent of Arizona Reading First students.

At the beginning of year on NWF scores, the gap in scores between "at risk" and "some risk" was about 13 points. The gap between the "some risk" and "low risk" groups was 19 points.

Looking at the average ORF scores for these groups in the spring, it was evident that those who started the year "at risk" had the lowest average ORF scores and did not catch-up to the average score for the middle or highest group. Table 5.6 presents these results.

Table 5.6
All RF: NWF Beginning of Year and ORF End of Year Mean Scores

Category	NWF Begin Mean	NWF Begin % of Total N	ORF End Mean	ORF End % of Total N
At risk	3.86	43.3%	11.08	29.7%
Some risk	17.74	23.2%	28.72	29.4%
Low risk	38.30	33.4%	66.70	40.9%
Total	18.60	100.0%	39.02	100.0%

The average fluency of the fall "at risk" group in the spring was 23.6 wcpm. This was about 15 points below the average ORF score of those starting the year in the "some risk" group; their average was 38.1 wcpm. Those students who started in the "low risk" group had an average fluency score of 59.6 wcpm, or 21 points higher than the some risk group.

Table 5.7
1st Grade: ORF End of Year Mean Score by NWF Beginning of Year Category

NWF Beginning Category	ORF End Mean	NWF Begin % of Total N	N	Std. Deviation
At risk	23.64	43.4%	2492	19.259
Some risk	38.09	23.2%	1336	22.059
Low risk	59.62	33.4%	1920	28.543
Total	39.02	100.0%	5748	28.114

The individual schools' mean scores were compared to the overall averages to determine the degree to which the overall pattern held across schools. In general, schools trends broke down as follows:

1. In 36 (57%) of the schools, the group "at risk" in the beginning of the year had an average end of year ORF score that was less than 15 points below the "some risk" group. This suggests that the schools may have closed the gap slightly and had some success in moving their lowest students ahead at a faster pace. Another possible interpretation, however, is that the students in the "some risk" group made lower-than-expected progress, so that the gap between the two narrowed not because the "at risk" group gained but because the "some risk" group did not gain enough. Schools in this group need to individually analyze their school's data to determine what explains a narrowing of the gap.

   The schools in this group were Andulacia, Challenger, Curry, Desert Sonora, Desert View, Gwyneth Hamm, H.L. Suverkrup, Holmes, Ironwood, J.B.Sutton, J.R.Davis, Lafe Nelson, Laird, LePera, Lynn-Uriquides, Menlo Park, Mitchell, Mountain View, Orange Grove, P.H.Gonzales, P.T.Coe, Palmcroft, Red Mesa, Robert Bracker,

Roosevelt-Mesa, Roosevelt-Yuma, Round Rock, Salida Del Sol, Stanfield, Sunland, Thomas J. Pappas, West, Westwood, Whittier, Willcox, and Wm.C.Clark.

2. In 27 schools, the group "at risk" in the fall according to the NWF had an average ORF score that was more than 15 points below the average of the "some risk" group, suggesting that these schools may need to pay particular attention to the instruction of their lowest performing first-grade students to ensure that they do not fall further behind their peers.

   The schools in this category were A.J.Mitchell, C.J.Jorgensen, C.E.Rose, Craycroft, Davidson, Drexel, Evans, Glendale American, Hawthorne, Lowell, MacLennan, Mesquite, M.O.Bush, Pendergast, Pueblo, Pueblo Gardens, Rainbow Valley, Roberts, Sevilla, Shaw Butte, Southwest, Summit View, T.G.Barr, Tierra del Sol, Valley Horizon, Westwind, and Whitman.

CHAPTER 6
ASSESSMENTS III: AIMS AND STANFORD 9 READING RESULTS

Highlights

Reading Assessments Summary for Third Grade: DIBELS, AIMS and Stanford 9

- All Arizona's third graders overall rated at higher levels on all three student assessments (DIBELS, AIMS and Stanford 9) than did the Reading First or Comparison school cohorts. Further, the RF and Comparison school groups had similar scores to each other (within a few points) across all three tests. These results are not surprising since the RF and Comparison schools were selected precisely on the basis of the number of students who were "at risk" in reading and the schools had high percentages of students who qualified for the free and reduced lunch program.

- There was no difference apparent to show that the RF students performed better than other students at the end of third grade following the first year of implementation. Thus, at this point, it is hard to see the positive impact of RF on the test scores of third grade students. The RF program was new to the schools and students in 2003-2004. Any "gap" in reading knowledge between where they were in their learning and the standard already would have existed for the third grade students, and have been very hard (albeit not impossible) to be closed in the first year of a new reading program. Further, there was some evidence to suggest that although new reading curricula were implemented in the first year, the interventions that would specifically target low students were not in place.

- In the next years of implementation, AIMS, Stanford 9 and DIEBLS results should be able to reveal more about the success of the RF programs, and the progress of individual students. As the students have successively more years of the science-based reading curricula and appropriate interventions, it is anticipated that improvements in test scores will accompany these additional years reading efforts.

AIMS Third Grade Reading Scores

- On the AIMS reading test, the third graders in Arizona had an overall lower weighted mean score in 2004 than in 2003.

- The mean scores of third graders in 2004 in the RF (505) and Comparison schools (507) continued to be lower than the mean score of third graders in the state (519).

- From 2003 to 2004, the percentage of the Arizona third graders who "Meet" or "Exceed" the standard *decreased* by 5 percent, in contrast to a decrease of 5.8 percent and 11.5 percent for third graders in the RF and Comparison schools, respectively. The goal is to increase those who "Meet" or "Exceed" the standard.

- In the "Falls Far Below" proficiency level category, it is always hoped that the percentage of students in this bottom category will decrease. However, between 2003 and 2004, RF third graders experienced a 6.9 percent increase and the Comparison school group saw an even

larger 11.6 upswing in the "far below" the standard category; third graders in the state experienced also an increase in this category although by only 4.0 percent.

Stanford 9 Second and Third Grade Reading Percentile Ranks

- On the Stanford 9 reading test, the mean percentile score of the RF second graders increased by 2 points in 2004 over 2003, whereas the mean score of second graders in the Comparison schools decreased by 2 points.

- The mean percentile scores of the RF and Comparison school third graders decreased slightly (by less than one point) from 2003 to 2004, but was still lower than the state average.

- For both second and third graders in all Arizona schools, the mean percentile scores on the Stanford 9 reading test did not change from 2003 to 2004 and were 11 to 15 points higher than the scores for the RF schools for both years.

- The second and third grade RF percentile scores were two to six points higher than the Comparison groups' scores in both 2003 and 2004.

Introduction

The results of the analysis of the AIMS and Stanford 9 tests are described in this section. The AIMS test includes three subject areas, reading, writing, and mathematics, which are administered to all students in grades 3, 5, 8 and 10 in the spring of the school year. For purposes here, only the third grade reading scores are used in the AIMS proficiency level analysis of the Reading First (RF) and Comparison schools. The proficiency levels include: "Falls Far Below the Standard" (FFB), "Approaches the Standard" (A)," "Meets the Standard" (M) and "Exceeds the Standard" (E).

The Stanford 9 is a standardized, norm-referenced test. In this section, the analysis of the Stanford 9 focuses on the second and third grade reading scores. The data are reported as percentile ranks for the school level. A percentile rank reflects the typical student's performance at the school compared to the norming group for that grade and subject area. Thus, if the school score is 39, it means that the average student at this school scored better than 39% of the students in the 1995 norming group. Schools with rans reported near the 50th percentile indicate that the typical student performance on the test is about average when compared with other students of the same grade level. Higher percentile ranks reflect better performance.

Several analyses of the reading tests are presented. The first analysis looks at the percentage distribution of third graders in the four AIMS proficiency level categories. These scores are compared for spring 2003 (2002-2003 year) to spring 2004 (2003-2004 year) data, and for the Reading First compared to the Comparison group of sixteen schools. Breakouts are also discussed for the percentage of third graders who fell in the lowest proficiency level ("Falls Far Below" standard) and highest levels ("Meets" or "Exceeds" the standards) between 2003 and 2004. Another analysis compares the percentage of schools that exhibited change in the scores on the third grade AIMS and Stanford-9 tests during the school year and the direction of the change. Finally, the analysis compares the scores of students in the sixty-three RF and sixteen comparison schools on the AIMS mean scale scores and the difference in scores in 2003 and 2004; and the second and third grade Stanford 9 mean reading percentile scores for the two test periods.

AIMS Proficiency Level Analysis

Table 6.1 displays the percent distribution of RF and Comparison third graders in the four AIMS proficiency level categories for the 2003 and 2004 spring tests.

The results in Table 6.1 indicate a negative skew (i.e., the tail is to the left) in that the majority of the third graders in the schools scored in the "Meets" and "Exceeds" categories or toward the higher end of the scale. In 2003 and 2004 respectively, 77 percent and 71 percent of the third graders in Arizona met or exceeded the standard. In 2003, 64 percent of the RF students met or exceeded the standards on the third grade AIMS reading test; the comparison schools did slightly better in that 66 percent met or exceeded the standard. In 2004, 58 percent and 54 percent of the third graders in the RF and Comparison schools respectively "Meet" and "Exceed" the standards on the AIMS reading test.

Table 6.1
Distribution of 3rd Graders in the Four Proficiency Levels

	Spring 2003				Spring 2004			
3rd grade AIMS Reading s	**% Falls Far Below**	**% Approach**	**% Meet**	**% Exceed**	**% Falls Far Below**	**% Approach**	**% Meet**	**% Exceed**
Reading First Schools	13.8	22.3	52.4	11.6	20.7	21.1	44.2	13.9
Comparison Schools	10.6	23.6	54.5	11.5	22.2	23.5	40.9	13.3
All Arizona Schools	8.0	16.0	56.0	21.0	12.0	18.0	46.0	25.0

The percentage of third graders in the RF and Comparison schools that scored in the lower categories (FFB and A) may be a cause for concern for the principals. The third graders in the RF and Comparison schools exhibited similar patterns of change. More specifically for all Arizona students, the percent of third graders who exceeded the standard in 2004 increased, while the percentage of students who were in the middle categories ("Approaches" and "Meets" standards) declined in 2004 from 2003.

The patterns of change reported by the schools reflect both positive and negative results in the high and low ends of the AIMS proficiency scale, respectively. In the next set of tables, the analysis looks at the two extreme levels of the AIMS scale, i.e., "Falls Far Below" on the low end and "Meets" or "Exceeds" the standards, on the high end of the scale.

"Falls Far Below" & "Approaches" Category
Table 6.2 displays the percentage of third graders in the "Falls Far Below" the standard category in the RF schools and Comparison schools, as well as the third graders in all Arizona schools, and the differences between 2003 and 2004. In general, the percentage of third graders who "Fall Far Below" the standard for both the RF and Comparison schools increased between the spring 2003 and 2004 test periods; this meant that a larger percentage of students were in the lowest category. It is also clear that the third graders in these schools did not do as well as third graders in Arizona overall. For students in RF schools, 13.8 percent fell far below the standards in 2003, while 21 percent fell far below in 2004, a 7 percent difference. For third graders in the comparison schools, 11 percent fell far below standard in 2003. It increased to 22 percent in 2004, or by 11 percent. Hence, more third graders than a year ago were categorized as "Falls Far Below" the standards in both the RF and comparison schools.

By contrast, the thirds graders in the comparison schools did not do as well as the third graders in the RF schools. The results showed that the third graders in the RF and Comparison schools ranked below the reading proficiency levels of students in Arizona schools as measured by the AIMS test. In the RF schools, the results suggested that the third grade reading program had not yet made an impact on the students most in need of the reading skills, i.e., those in the "Fall Far Below" standard category. However, since the reading programs were implemented in the schools only beginning in the fall of 2003, a major impact on the AIMS reading scores would not

have been expected for the spring of 2004. The new reading curricula were not fully implemented in the first year and the interventions needed by the lowest students to bridge the gap were not in place. This same finding was indicated by the third grade DIBELS scores, these "low" students were the least impacted by the Reading First program.

Table 6.2
Percent of Students in "Falls Far Below" Category 2003 and 2004*

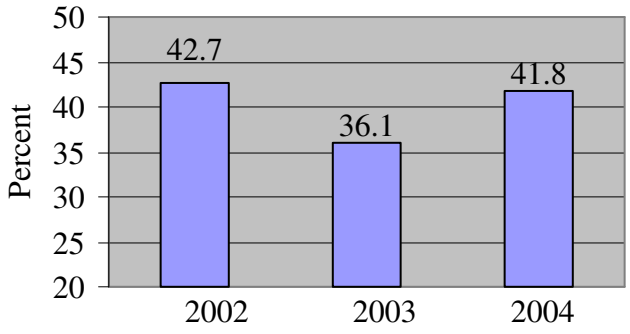
Category	% Spring 2003*	% Spring 2004*	% Difference
Reading First Schools	13.8	20.7	6.9
Comparison Schools	10.6	22.2	11.6
All Arizona Schools	8.0	12.0	4.0

*Since FFB is the low category, lower percentages are desirable; the percent difference should decrease to show improvement.

Those students "not meeting the standard" are a combination of the "Falls Far Below" and the "Approaches" categories. The results shown in Figure 6.1 indicated that the percentages of RF students not meeting the AIMS standards have not shown any distinct pattern over the course of three years. During the base year of 2001-2002, 42.7 percent of the students fit this category. That figure dropped to 36.1 in 2002-2003, but shot back up to 41.8 percent by 2003-2004. Consequently, the net decline of third graders not meeting the AIMS standard dropped by less than 1 percent over the three span.

Figure 6.1

RF 3rd Grade Students "Not Meeting" AIMs



Note: "Not Meeting "is a combination of "Falls Far Below" and "Approaches" categories

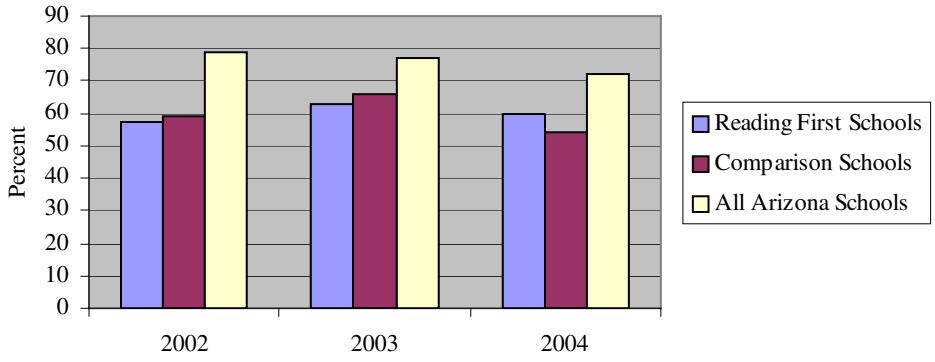
"Meets" or "Exceeds" Standard
Table 6.3 presents the percentage of third graders who "Meet" or "Exceed" the standard, and the difference between the spring 2003 and 2004 testing periods.

The results indicated that the percentages of RF students who "Meet" or "Exceed" the standard declined slightly to 58.2 percent in the spring 2004 test, from 64.0 percent in 2003. The third graders in the Comparison schools also exhibited a similar pattern, although the percent reduction was even greater than that observed for the RF schools. In general, the percentage of third graders who "Meet" or "Exceed" the standard dropped for all students on the AIMS test in 2004 from 2003. The 2004 scores for the RF group were just above their 2002 level; RF students had not dropped as much in total over the two years as had the Comparison group or all Arizona schools.

Table 6.3
Percent of Students in the "Meet" and "Exceed" Categories: 2003 and 2004

Category	% 2002	% 2003	% 2004	% Difference 2003 to 2004
Reading First Schools	57.3	64.0	58.2	-5.8
Comparison Schools	59.3	68.8	54.3	-11.5
All Arizona Schools	79.0	77.0	72.0	-5.0

Figure 6.2
3rd Grade Students "Meet" or "Exceed" AIMs



It is not surprising that the percentage of Arizona's third graders who "Meet" or "Exceed" the standard is higher than third graders in either the RF schools or the Comparison schools, since the latter schools were selected precisely on the basis of the number of students who were "at risk" in reading. Basically, all students progressed in the *wrong* direction showing a decline in

those who met or exceed the standard from 2003 to 2004; the percent of RF students in the meet or exceed groups declined less than their Comparison counterparts.

AIMS Gains and Losses

Table 6.4 shows the percent of the RF schools in which the FFB's increased or decreased between 2003 and 2004. The objective here is to examine whether the RF schools were able to raise students who "Fall Far Below" standard up to the next proficiency level. The results, however, indicate that the RF schools have yet to succeed in this objective. In 2004, 73 percent of the RF schools had more third graders who fell in the FFB category than in 2003 (more third graders did not do as well). By contrast, only 21 percent of the RF schools' third graders saw a decrease to the FFB category (the third graders did better) in 2004. So, the RF third graders in the FFB category did not improve on the AIMS test in 2004 after a year's exposure to the new RF reading curriculum.

Table 6.4
AIMS Test Scores Changes from
2003 to 2004 –"Falls Far Below" Category

Change from 2003 to 2004	N of Schools	% of School
Improved	13	20.6
Worse	46	73.0
No Change	4	6.4
Total	63	100.0

Stanford 9 Second and Third Grade Reading Test Percentile Rank Scores

It is important to remember that the Stanford 9 scores are reported as percentiles; if the school score is 39, it means that the average student at this school scored better than 39% of the students in the 1995 norming group

Table 6.5
Stanford 9 Reading Scores 2nd and 3rd Grades 2003 and 2004

Stanford 9	2nd Grade		3rd Grade	
2nd & 3rd Grade Reading Percentile Rank Scores	Spring 2003	Spring 2004	Spring 2003	Spring 2004
Reading First Schools	44.2	46.0	39.6	39.7
Comparison Schools	42.1	40.4	35.5	36.3
All Arizona Schools	57.0	57.0	54.0	54.0

For practical purposes, the Stanford 9 reading scores for all Arizona schools, for the RF schools and the Comparison schools did not change in the 2004 test from the 2003 test. The RF second and third grade rank scores on the Stanford 9 moved slightly in the positive direction for 2004 from 2003; the second grade rank improved approximately two points, whereas the third grade ranking moved only 0.2 points. However, none of the RF scores moved above the mid-point rank of 50 percentile.

For second grade students in the Comparison group, however, the slight change for 2004 was negative (less than two points); the third grade change was less than one point in the positive direction. The Comparison group, like the RF schools, did not rank above the 50 percentile mid-point. The RF schools did have higher rankings at both the second and third grade level then did the Comparison schools.

In all Arizona schools (the RF and Comparison schools are included in the all schools totals), the average second grade rank was 57 in 2003 and 2004, 11 points higher than the RF schools. The average third grade rank was 54 in 2003 and 2004, 14 points higher than the 2004 RF students.

Stanford 9 Second and Third Grade Reading Test: Schools' Gains and Losses

Table 6.6 shows the percentage of the RF schools that showed gains or losses on the Stanford 9, second and third grade reading tests from the 2003 to the 2004 test. Again it can be seen that the RF schools did not improve. A few more schools reported gains than losses, both in second and third grades across the 2003 to 2004 period. Of the RF schools for second graders, the findings indicated that 50.9 percent of the schools reported that they had gains in scores from 2003 to 2004, while 43.6 percent of schools showed a decrease in the percentile score. Of the RF schools for third graders, the findings indicated that 53.6 percent of the schools reported that they had

gains in scores from 2003 to 2004, while 46.4 percent of schools showed a decrease in the percentile score.

Table 6.6
RF Stanford-9, 2nd and 3rd Grade Reading Scores:
% Difference from 2003 to 2004

Difference from 2003 to 2004	2nd Grade (n=55) %	3rd Grade (n= 57) %
Gained	50.9	53.6
Loss	43.6	46.4
No Change	5.5	0.0
	99.0	100.0

In Table 6.7, the percentage of RF schools that improved or not between 2003 and 2004 on the second grade reading test is compared with the Comparison schools. The RF schools did about the same (49.1% to 50% respectively) as the Comparison schools as indicated by the percentage that gained in percentile ranking on the reading test scores between the 2003 and 2004 testing periods. However, if the percentage of "no change" is added to the percentage of schools that did "better" in the spring (49.1% + 7.3% = 56.4%), then the RF schools can be said to have moved slightly in a more positive direction than did the Comparison schools.

Table 6.7
RF and Comparison Stanford 9 2nd Grade Reading Scores:
% Difference from 2003 to 2004

Difference from 2003 to 2004	RF Schools (n =55) %	Comparison Schools (n = 12) %
Gained	49.1	50.0
Loss	43.6	50.0
No Change	7.3	0.0
	100%	100%

The data in Table 6.8 show the percentage of RF schools and Comparison schools that gained or lost in the third grade reading test scores between the 2003 and 2004 tests. Indeed, 53.6 percent of the RF schools showed gains in test scores from 2003 to 2004. In contrast, only 40 percent of the comparison schools reported gains in reading scores. Indeed, 60 percent of the Comparison schools had lower scores in 2004 than in 2003, as compared with 46.4 percent of the RF schools.

A higher percentage of RF schools did "better" on the test in 2004 than the Comparison schools. Although the gains are modest, they are meaningful in that some RF schools that exposed students to the new reading program were able to show slight gains.

Table 6.8
RF and Comparison Stanford 9 3rd Grade Reading Scores:
% Difference from 2003 to 2004

Difference from 2003 to 2004	RF Schools (n=57) %	Comparison (n=10) %
Gained	53.6	40.0
Loss	46.4	60.0
	100.0	100

Progress on the AIMS and Stanford 9 Reading Tests

Each year, progress on student reading performance on the third grade AIMS and second and third grade SAT 9 are used as indicators of gains made among students in all participating schools in Arizona.

Table 6.9
Reading Tests: 2003 and 2004

	AIMS 3rd Grade Mean Scores			SAT 9 2nd Grade Percentile Rank			SAT 9 3rd Grade Percentile Rank		
	2003	2004	Diff	2003	2004	Diff	2003	2004	Diff
Reading First Schools	510.8	505.2	-5.6	44.2	46.0	1.8	39.6	39.7	0.1
Comparison Schools	512.9	507.4	-5.5	42.1	40.4	-1.7	35.5	36.3	0.8
All AZ Schools	523.0	519.0	-4.0	57.0	57.0	0.0	54.0	54.0	0.0

Note: For analysis purposes, AIMS weighted means were computed for the RF and comparison schools in order to factor in the size of the student population.

The results displayed in Table 6.9 show that the AIMS third grade weighted mean test scores were lower in 2004 than in 2003 for students in the RF, Comparison schools, and Arizona. In fact, the mean scores for students in both RF and Comparison schools declined by nearly six points in 2004 from 2003. For the RF students, the mean score declined from 510.8 to 505.2, and for the students in the Comparison schools the mean score declined from 512.9 to 507.4. The RF schools' mean score was two points below the Comparison schools' mean scores for both years. For all Arizona students, the mean score declined by four points (from 523 to 519). The mean scores for the RF and Comparison schools were still 14 to 12 points respectively, *below* the state average in 2004, as measured by the third grade AIMS reading test.

For practical purposes, it can be said that the Stanford 9 reading test mean percentile ranks of the RF and Comparison group schools did not change in the 2004 test from the 2003 test. The RF second graders did show a minimal improvement with the second grade percentile rank changing by two points from 2003 to 2004. The third grade percentile score held steady at just below 40. For second grade schools in the Comparison group, however, the change during the year was negative, and the third grade change was less than one point in the positive direction. The RF schools exhibited slightly higher percentile rankings on the second and third grade Stanford 9 reading tests than did the second and third grade Comparison schools. There is a gap (from 11 to 18 points) in the mean percentile ranks between students in all Arizona's schools and schools in the RF and Comparison groups. The RF and Comparison schools are below the state average on the Stanford 9 reading tests.

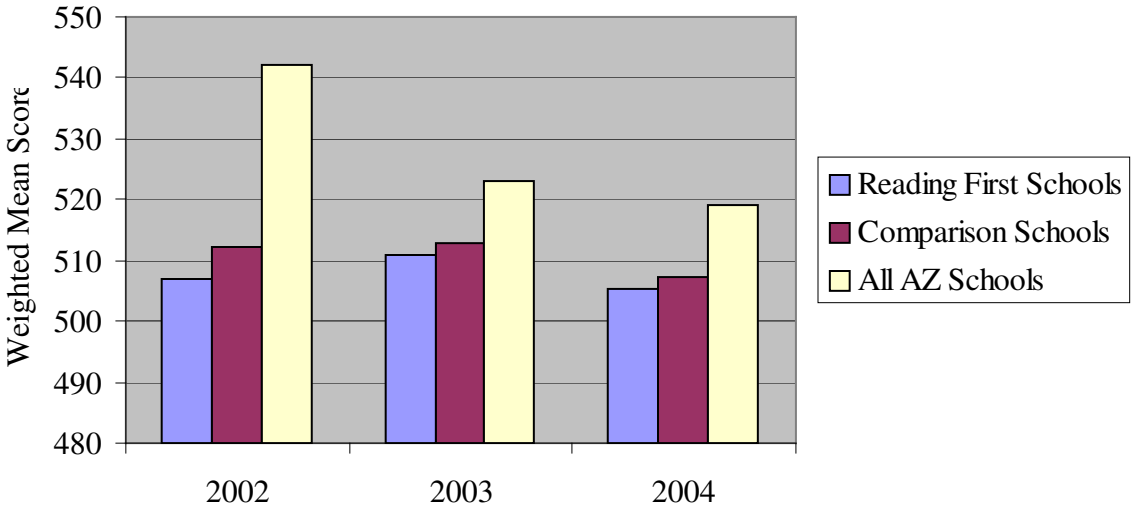
When compared across the past three years, the AIMS weighted means scores varied little for each of the groups of schools. While the 2003 to 2004 scores had declined slightly for each of

the groups, the RF and the Comparison groups had shown slight increases from 2002 to 2003. The biggest two year differences were that all Arizona schools declined by 23 points, and the Comparison group gained 14 points.

Table 6.10
Reading Tests: 2002, 2003 and 2004

	AIMS 3rd Grade Weighted Mean Scores		
	2002	**2003**	**2004**
Reading First Schools	507.1	510.8	505.2
Comparison Schools	512.2	512.9	507.4
All AZ Schools	542.0	523.0	519.0

Figure 6.3
AIMS 3rd Grade Reading Mean Scores



Reading Tests Summary for Third Grade: DIBELS, AIMS and Stanford 9

The third grade is the only level that crosses the three tests. However, it is difficult to cross-reference these tests, as the categories by which students are classified have very different meanings. The DIBLES end of year ORF instructional support recommendation categories showed that only 30.5 percent of students had met the "benchmark;" on AIMS, 44.2 percent were "Met" and another 13.5 percent had "Exceeded" the standard. Of greatest importance were the large percent of students who were *below* the standard: on DIBELS, the large percent of students in the intensive (36.7%) and strategic (32.8%) categories were not at the standard, while the AIMS showed 20.7 percent "Falling Far Behind" and 21.1 percent "Approaching" the standard. On the third grade Stanford 9, the average ranking at the 40th percentile for the RF schools also indicates that many students are below the expected norm.

Figure 6.4
Reading First 3rd Grade End of Year ORF

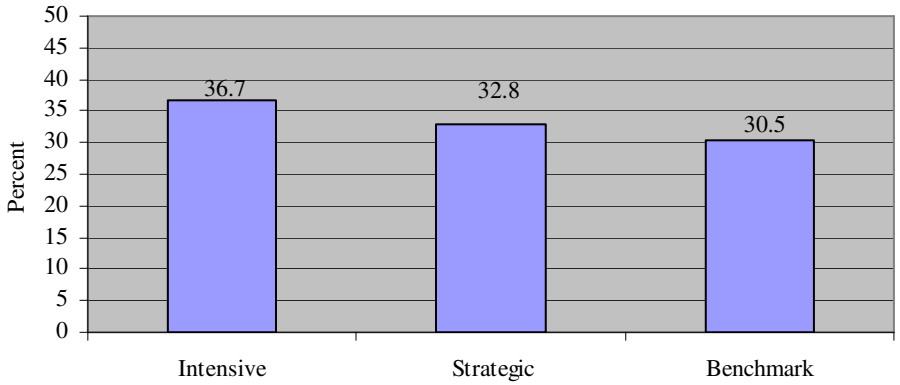
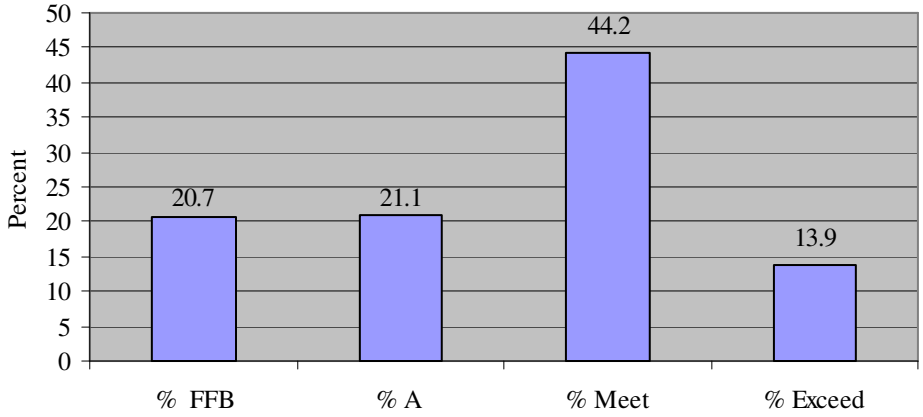


Figure 6.5
2004 AIMS 3rd Grade Reading



Another examination of third grade reading assessment scores looks at the patterns across the RF, Comparison and all Arizona schools groups (Figures 6.6, 6.7 and 6.8). The RF group comprises the 63 schools of the 2003-2004 first year of implementation. The Comparison group combines all 16 of these schools (in the initial DIBELS analysis in Chapters 4 and 5, the 16 schools were divided into two groups). All Arizona schools also include those in the RF and Comparison groups (DIBELS scores were not available for the all AZ schools group).

All Arizona's third graders overall rated at higher levels on all three measures than did the Reading First or Comparison school cohorts. On the AIMS and Stanford 9 tests, the scores/ranks for all Arizona Schools were the highest; scores/ranks for the RF and Comparison schools were both substantially lower than the Arizona schools as a whole.

For the third grade, the RF and Comparison school groups had similar scores to each other (within a few points) across all three tests. On AIMS, the RF scores were slightly lower than the Comparison group; on the Stanford 9, the RF ranks were slightly higher than the Comparison schools; and for the DIBELS scores, the two groups were basically the same. The "no gain" by third graders in the RF program on their DIBELS score that would have predicted little to no impact on the standardized testing.

At this point, there was no difference apparent based on third grade student assessment scores to show that the RF students improved more than other students at the end of third grade. This finding could have been expected since the RF and Comparison schools were selected precisely on the basis of the number of students who were "at risk" in reading, and the large percentage of students who qualified for the free and reduced lunch program.

The RF program was new to the schools and students in 2003-2004. Third grade students would not have had the basics instilled in them by the kindergarten, first and second grade years of a science-based reading curriculum. Any "gap" in reading knowledge between where they were in their learning and the standard already would have existed for the third grade students, and have been very hard (albeit not impossible) to be closed in the first year of a new reading program. Further, there was some evidence to suggest that although new reading curricula were implemented in the first year, the interventions that would specifically target low students were not in place. Thus, although they had an improved curriculum in the third grade, it did not seem to be "enough" by itself to raise their test scores over what was found to be the "norm" for an "at risk" student group.

In the next years of implementation, AIMS, Stanford 9 and DIEBLS results should be able to reveal more about the success of the RF programs, and the progress of individual students. As the students have successively more years of the science-based reading curricula and appropriate interventions, it is anticipated that improvements in test scores will accompany these additional years reading efforts.

Figure 6.6
2004 AIMS 3rd Grade Scores

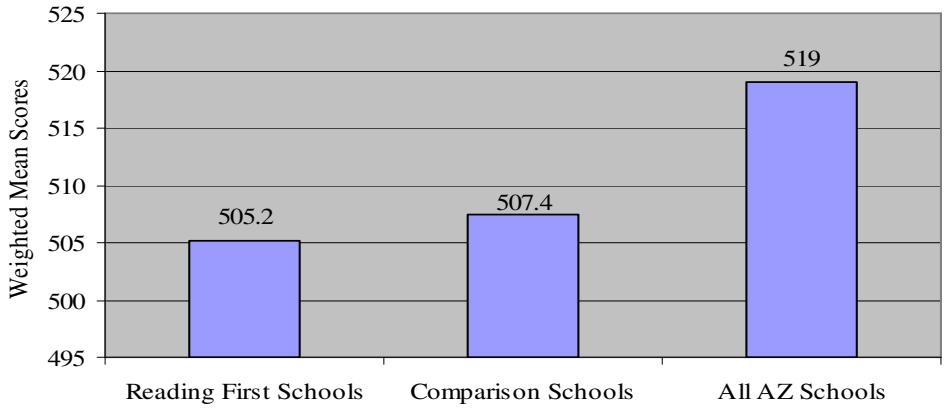


Figure 6.7
2004 Stanford 9 3rd Grade Ranks

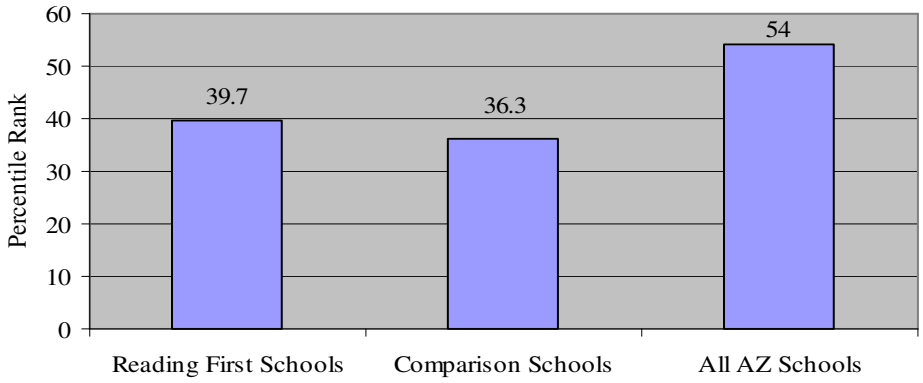
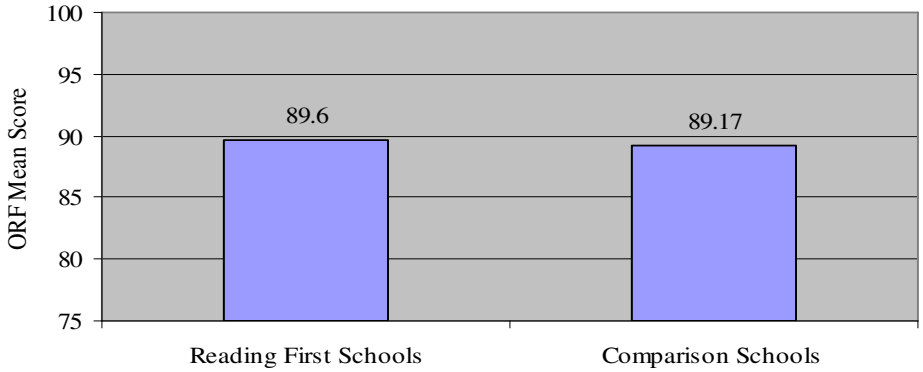


Figure 6.8
2004 DIBELS 3rd Grade End of Year ORF



CHAPTER 7
PROFESSIONAL DEVELOPMENT

Highlights

- DIBELS Administration training had a sizeable impact on participants' understanding and ability to use the assessment. Over 85% report they were "confident" to "very confident" in their ability to interpret DIBELS.

- The LETRS Institutes made a positive contribution to participants' learning. The majority said the Institute was also relevant to immediate work and enhanced their ability to use SBRR.

- The 19 item knowledge questions regarding the five essential reading components had mean scores increase 10 to 15 percent from the pretest to the after the training post-test. In contrast, the scores on the follow-up spring survey resulted in a decrease of 10 to 12 percent when compared to the posttest immediately after the training.

- Schools and districts at further distances from Phoenix were much more likely to feel that the distance required to travel for training events was onerous in terms of time and money, particularly for the monthly meetings.

- Principals spoke positively about the professional development provided to them by the state. About half of them said that these trainings had caused them to gain a better general understanding of the Reading First program and a specific understanding of the five essential reading components.

- When asked in interviews what additional training they would like, some principals suggested additional training in motivating staff, "working with resistant teachers" and "building buy-in." Others suggested further training in how to use data to drive instruction, specifically "how that would look on a day to day basis."

- Coaches overwhelmingly indicated that the training provided to them by the ADE had been very helpful. Most coaches cited the monthly meetings, with the vast majority of comments being very positive regarding their usefulness. Conversely, a small number of coaches and specialists complained that some of the monthly meetings were redundant or not applicable to their position.

- About half of coaches said that they would like additional training in coaching methods. As many were new to the coaching role, they requested more definition of their role and assistance in implementing it.

- Almost all assessment coordinators had positive feedback regarding the professional development that they had received through the state; they found it "useful," "excellent" and were "amazed with the quality."

- Assessment coordinators requested more training in various aspects of working with data. The most frequent request was for help with data presentation, including: communicating data to teachers, helping teachers "own" their data, and instruction in how to read the "new" DIBELS graphs.

- There was weak or little evidence indicating substantial training in the use of assessment data or strategies for ELL students. The next year, ADE should make these two topics a priority.

- Teaching staff indicated that, of all professional development that they received, they were most impressed by the professional development provided at the school-level by the reading coach.

- Vocabulary, comprehension, and fluency were the top three components pointed out as areas in which a need was indicated for additional training for teachers. Some also expressed an interest in receiving additional training in classroom management, student engagement, using data and assessment to guide instruction, and working with English language learners.

- Many of the district representatives interviewed specifically commented on the "great training." For the most part trainings were rated as very helpful, and "the training has been very good because it gives a similar focus" to schools.

- In interviews, principals and coaches overwhelmingly praised the ADE for the support it provided them in the first year of the grant.

- County Reading Specialists provided a wide range of services based on the schools' needs and were largely viewed as very supportive, although a few schools had less positive experiences. They played an important role in supporting reading assessment by assisting with DIBELS training, administration, and data interpretation.

- Six of 13 reading specialists said they were "well" or "very well trained/prepared" to provide expert service to the schools. Seven were less forceful in their response and indicated that they were "somewhat well trained" for the reading specialist role.

- Reading specialists also ranked their three greatest challenges. The "lack of teacher buy-in to the Reading First goals" ranked first on the list of seven items checked by the specialists. "Lack of time to provide all the services requested or needed" ranked second, followed by the "location of or convenient access to the schools," in third place.

Introduction

A major component of Arizona Reading First is the provision of professional development to principals, reading coaches, assessment coordinators, county reading specialists and staff. This is one means to a achieve Arizona's Reading First objective of implementing a research-based comprehensive instructional program through capacity building. Therefore, the ultimate purpose of professional development is to influence the content and quality of reading instruction at the classroom level.

This section of the report summarizes professional development activities in the first year of Arizona Reading First and reviews feedback from participants about the quality, relevance, and utility of professional development.

DIBELS Administration Training Feedback Surveys

In August-September 2003, the Arizona Department of Education (ADE) held two training workshops to guide Local Education Agency (LEA) assessment teams in the use and interpretation of the *Dynamic Indicators of Basic Early Reading Skills* (DIBELS) assessment. The first workshop focused on administration of the DIBELS; after this workshop participants were able to return to their schools to administer the assessment. The second workshop, held two weeks later, focused on interpretation of DIBELS data. Participants were invited to bring their schools' results to this training to learn how to interpret these data.

Feedback surveys were collected from 473 participants in the DIBELS Administration session and 228 participants from the DIBELS Interpretation session. Participants were primarily reading coaches, assessment specialists, principals, and other school and district administrators.

Participants were asked to rate their understanding of the DIBELS on a scale of zero to six (with a zero denoting no prior understanding and a six denoting a thorough understanding) at three points: (1) prior to all Reading First training, (2) prior to the DIBELS Administration training, and (3) after the DIBELS Administration training. The results, presented in Table 7.1, indicated that participants had relatively low levels of understanding of the DIBELS prior to Reading First and that other Reading First trainings prior to this had not substantially increased their understanding of the DIBELS. However, the results showed that this DIBELS Administration training had a large positive impact on participants' understanding of the assessment, increasing the average rating of understanding from 1.76 to 4.70 on the seven-point scale.

Table 7.1
Self-Reported Understanding of DIBELS Assessment Before and After Training

	N	Mean*	Standard Deviation
Prior to all Reading First training	468	1.04	1.446
Prior to DIBELS Administration training	467	1.76	1.565
After DIBELS Administration training	468	4.70	0.910

* Mean is based on a seven-point scale where six was highest level of understanding

Participants were asked to rate their ability to interpret DIBELS data on a scale of zero (low) to six (high) at two points: (1) prior to the training, and (2) after the training. The results, presented in Tables 7.2 and 7.3, indicated that the training increased participants' self-reported ability to interpret DIBELS data from an average rating of 3.47 to 4.70. In terms of numbers, those rating their ability as "very good" (5) or "very strong" (6) went from 21.0% prior to the training to 57.5% after the training (and an additional 30% rated their ability as "good" (4) after the training).

Table 7.2
Self-Reported Ability to Interpret DIBELS Data Before and After Training

	N	Mean*	Standard Deviation
Prior to today's training	208	3.47	1.344
After today's training	223	4.70	1.011

* Mean is based on a seven-point scale where six was highest level of understanding

Table 7.3
Self-Reported Ability to Interpret DIBELS Data Before and After Training

Rating of Very Good (5) or Very Strong (6) Ability	*N = 228* Frequency	Percent
Prior to today's training	48	21.0%
After today's training	131	57.5%

Participants were also asked to rate their confidence in interpreting the DIBELS data for different grade levels on a scale of zero (low) to six (high). These results, presented in Table 7.4, indicated that by the end of the training, the majority of respondents were confident with interpreting the DIBELS at all grade levels. It should be noted that participants had very similar levels of comfort across grade levels; the consistency of ratings across grade levels in these results suggests that grade level does not have an effect on comfort with interpreting the assessment. Over 85% report they were "confident" to "very confident" (rating 4, 5, or 6) in their ability to interpret DIBELS.

Table 7.4
Self-Reported Confidence in Interpreting DIBELS Post-Training

Confidence interpreting the DIBELS	K (n=209)	1 (n=442)	2 (n=441)	3 (n=439)
0 (Not Confident)	0.0%	0.0%	0.0%	0.0%
1	0.5%	0.5%	0.5%	0.5%
2 (Somewhat Confident)	1.9%	1.9%	2.4%	3.9%
3	10.0%	9.1%	9.1%	8.8%
4 (Confident)	32.1%	31.6%	31.3%	30.9%
5	31.1%	31.6%	30.8%	29.9%
6 (Very Confident)	24.4%	25.4%	26.0%	26.0%

LETRS Institutes Fall 2003

In fall 2003, the Arizona Department of Education held the second set of Language Essentials for Teachers of Reading and Spelling (LETRS) Institutes. This cycle consisted of the nine modules in three books: Book One, Foundations of Reading Instruction, Modules 1, 2, and 3 in September; Book Two, Vocabulary, Fluency and Comprehension, Modules 4, 5, and 6 in November; and Book Three, Teaching and Assessing Beginning Reading and Spelling, Modules 7, 8, and 9 in December.

This fall cycle was specifically for Reading First participants and included LEA related reading coaches, reading specialists, and curriculum coordinators (28), as well as ADE staff and state reading specialists (16), and a few university/partner staff (4). The actual number who attended varied by Institute, and the number completing the surveys varied.

The LETRS feedback survey also included two ***content*** items that asked the respondents the extent to which the LETRS Institute (1) "enhanced their ability to use SBRR," and (2) was "relevant to [their] immediate work." In fact, more than two-thirds (69.4%) of the participants of the September Institute indicated that the institute "enhanced their ability," and 69.4 percent also said the Institute was also relevant to immediate work. The November and December Institute participants gave similar ratings to the "ability" item. Hence, approximately fifty-two percent said that the institutes "enhanced ability to use SBRR (51.6% and 51.9% respectively). Almost two-thirds, (64.3%) of the December participants said that the institute was "relevant, to a great extent, to immediate work" as compared with 54.8 percent of the November participants.

Table 7.5
Rating of LETRS fall 2003 Content

% Rating as "Superior"	September Mods 1, 2, 3	November Mods 4, 5, 6	December Mods 7, 8, 9	Total
CONTENT				
Enhanced Ability to Use SBBR	69.4%	51.6%	51.9%	58.5%
Relevant to Immediate Work	69.4%	54.8%	64.3%	63.2%

For the most part, the majority of the participants who attended the three LETRS Institutes gave "superior" ratings to the Institutes' format in terms of overall quality: means scores were 5.78 in September , 5.52 in November and 5.39 in December (0-6 scale with 6 highest); 63.2%overall rated the overall quality as "superior." A large majority of participants (average of 93.6%) viewed the LETRS sessions level of difficulty as "just right."

Knowledge and Attitudes

Knowledge Test
The data in Table 7.6 present the pretest, posttest and follow-up knowledge scores as measured by the 29-item survey. It can be seen that the mean scores of the key players increased 10 to 15 percent from the pretest to post-test (pretest before training and posttest after training). Indeed, in the long term, the scores on the follow-up in the spring still showed a slight increase in the scores from the pretest. However, the knowledge scores resulted in a decrease of 10 to 12 percent on the follow-up in the spring compared to the posttetst immediately after the training.

This is true for each group, which exhibited roughly the same decrease from the fall to the spring term. It can be seen that the coaches had a slightly greater reduction in knowledge due to higher baseline mean score. Although the key players taught and utilized the essential components of reading which were included in the knowledge test during the school year, the follow-up knowledge level as measured by mean scores for the three groups declined; for teachers and specialists, their scores were still above pre-academy levels (pre-academy teachers = 55.93, specialists = 58.44), although the coaches did drop to their pre-academy levels (67.82). It should be noted that changes in knowledge do not always translate into changes in practice. Although knowledge of SBRR as measured by the test decreased, this inability to answer a question correctly does not mean that the participants are ***not*** using effective instructional practices in the classroom. They may indeed be using these practices, but forgotten some of the nuances and language they learned in the academy. For better information about the degree to which knowledge translates into the classroom, observations in the classroom may serve this purpose.

Table 7.6
Pre, Post and Follow-Up Test Knowledge Scores

*Post & Follow-up "Matched"	Teachers (N=520)	Specialists (N=29)	Coaches (N=54)
Pretest mean score (SD)	55.93 (13.39)	58.44 (12.93)	67.82 (11.54)
Posttest mean score * (SD)	70.48 (12.45)	73.60 (12.03)	79.76 (13.27)
Follow up-mean score * (SD)	60.09 (12.76)	62.78 (14.29)	67.69 (13.27)
Mean difference (Follow up to Posttest)	10.39***	10.82***	12.07***

*** Significant at p<.001
Note: The data present knowledge scores as a percentage correct out of a possible 100 percent.

Results from the May 2004 Checklists

Introduction
The implementation checklist included eight questions about the content and audience for professional development. These items describe the existence of or participation in professional development but do not measure the quality of professional development offerings. As presented in Table 7.9, positive responses from CRSs about professional development questions were fairly high. CRSs indicated that almost all schools had fully participated in the AZ READS Summer Academy and in the training by the publisher of the core reading program.

In addition, the data indicated that all but two schools showed at least "some" evidence of ongoing professional development for all Reading First staff. Over half had "fully implemented" ongoing professional development in the core reading program and reported that the ongoing professional development had targeted the needs of staff and students. While evidence that schools received trainings in DIBELS, and the training for ELL and special education students had increased since February, these remained the areas most in need of development based upon the percent that had "no/scant" training in these areas.

Table 7.9
Implementation Checklist Items Related to Professional Development

	May 2004 Percent (n)			Change from February*
All K-3 teachers participated in the AZ READS summer academy.	90.5 (yes) (57)			-1 school
All K-3 teachers participated training by the publisher of their core reading program.	93.7 (yes) (59)			+ 4
	No/ Scant	**Some**	**Full**	
Ongoing professional development includes the principal, Reading Leadership Team, coach, assessment team, special education and other specialists, and K-3 classroom teachers.	3.2 (2)	22.2 (14)	74.6 (47)	+ 9 schools
Ongoing professional development targets the identified needs of staff and students.	4.8 (3)	36.5 (23)	58.7 (37)	+ 11
K-3 teachers participate in ongoing training in the use of the core reading program.	6.3 (4)	36.5 (23)	57.1 (36)	+ 10
Ongoing professional development is provided in research-based intervention strategies.	9.7 (6)	40.3 (25)	50.0 (31)	+ 15
Professional development in research-based strategies is provided for ELL and special education teachers.	12.7 (8)	23.8 (15)	63.5 (40)	+ 12
K-3 teachers received training in use of DIBELS assessment data.	14.3 (9)	23.8 (15)	61.9 (39)	+ 13

*Number of schools that increased from "no" or "scant evidence" to "some evidence" or "fully implemented" from February to May. See text for discussion of limitations.

Professional Development Experience.

Participants' reactions to professional development activities in the first year of the grant were obtained during site visits. In interviews, principals, reading coaches and assessment coordinators provided extensive qualitative feedback regarding the professional development that they had received through the State. Input from other school staff was solicited through focus groups held with the Reading Leadership Team members. Interviewees were asked to discuss the usefulness of the training they had received, provide a specific example of something that they had learned and used from a training, and suggest areas in which they would like additional training.

Professional Development vs. Teaching Experience
The pre- and follow-up surveys asked principals, teachers and specialists to indicate their degree of agreement with the statement: "While teachers' opportunities for professional growth are valuable, it is ultimately my/their years of experience that have taught them/me the most about being an effective teacher." The purpose here was to probe the attitudes of the participants regarding the value of professional development.
The responses are presented in Table 7.7

The follow-up data show that participants continue to place a higher value on their years of teaching experience, with 50.2 % who "agreed" or "strongly agreed" with this statement, up

from 45.4% at the pre-test. The majority of the respondents were teachers, for whom the "true test" of professional development is often its application in the classroom and their abilities to use and integrate their learning.

Table 7.7
Perceptions of the Relative Value of
Professional Development and Experience

Statement: While teachers' opportunities for professional growth are valuable, it is ultimately their/my years of experience that have taught them/me the most about being an effective teacher.		
	Pre-Test (n=563)	**Follow-up Test (n=554)**
Strongly Agree/Agree	45.4% (256)	50.2% (278)
Neutral	27.2% (153)	26.0% (144)
Strongly Disagree/Disagree	27.3% (154)	23.8% (132)

Professional Development Offerings from the State
Many professional development trainings were provided by the ADE to help schools use effective instructional assessment practices to improve reading. These trainings were provided at multiple levels and by a variety of trainers, mentors, and coaches.

A centerpiece of the training efforts in the first year of the grant was the 2003 Summer Academies. Over 1,400 teachers, specialists, principals, reading coaches, and assessment coordinators attended these four-day academies, which included both nationally recognized keynote speakers and a wide range of smaller breakout sessions targeting areas of specific interest and need.

The state also held specialized, trainings for coaches once a month and for principals every other month; these meetings were designed to give coaches and administrators the knowledge and tools they needed to solve problems at their schools. Training sessions were often facilitated by the state program staff but sometimes involved outside trainers as well. Table 7.8 lists the training sessions offered through the ADE in the first year of Arizona Reading First.

Table 7.8
Professional Development through the State

Event Description	Topics Included…	Invited attendees
2003 Reading First Summer Academies	Five essential components of reading instruction; overview of Reading First; leadership	Teachers and staff Principals Reading Coaches Assessment Coordinators
DIBELS Training and Follow-up Training	Administration of DIBELS ("Train the Trainer"); Analysis and interpretation of DIBELS data	Principals Reading Coaches Assessment Coordinators
LETRS Trainings (I, II, and III)	LETRS modules	District training representatives
County Reading Specialists' Training	Roles and Responsibilities, Working with the County Office, Assignments; Reporting with the Implementation Checklist, DIBELS Training Plan, Computer Support, Core Reading Programs; Principles of Cognitive Coaching; DIBELS; Building Leadership Capacity; Implementation Checklist Process and Questions	County Reading Specialists
Site-Based Reading Coaches' Training	Roles and Responsibilities, Adult Learning Theory, Coaching Theory and Strategies; Coaching Classroom Teachers; Observing Lessons and Giving Formative Feedback	Reading Coaches
	Using Assessment to Guide Intervention; Reviewing Benchmark Data, Analyzing DIBELS Data; Working with Grade-Level Teachers to "Own Their Data"; English Instruction for ELL students, Phonics and Comprehension for ELL students	Reading Coaches with Assessment Coordinators
Principals' Training	A Reading System that Increases Achievement; Observation Training and Walk-About Schedule; Planning Year 2 Budget; Professional Development Planning; "Let's Look at Our Data!"	Principals
Special LEA Training	Selection of Intervention and Supplemental Reading Programs	LEA Representatives

Feedback on State-Provided Professional Development

In interviews, principals, reading coaches and assessment coordinators provided extensive qualitative feedback regarding the professional development they received through the State. Input from other school staff was solicited through focus groups held with the Reading Leadership Team members. Interviewees were asked to discuss the usefulness of the training they had received, provide a specific example of something that they had learned and used from a training, and suggest areas in which they would like additional training.

Principals

Principals spoke positively about the professional development provided to them by the state. About half of them said that these trainings had caused them to gain a better general understanding of the Reading First program and a specific understanding of the five essential reading components. They also reported that they acquired skills to better monitor teacher implementation (e.g., conducting walkthroughs and completing evaluation forms) and better manage the grant.

> *Training in the five key components of Reading First included good examples of what instruction looks like and what strategies could be observed in grade level implementations. This training helped me monitor the implementation. (Principal)*

> *Huck Fritterer's meeting [about walk-throughs] was essential to helping me do my duties. Walk-throughs are one of the main duties of principals in the program. (Principal)*

About one third of principals said that they appreciated the opportunity that trainings provided for networking, collaboration, and data sharing with administrators at other schools. They liked how the grant "has given them unity and the ability to reach for a common goal."

> *It is useful networking and meeting other principals who are in the same boat and seeing how they got there, why they are there. It was helpful, encouraging, we celebrated each others success. It gave us resources to call each other. (Principal)*

Finally, several principals specifically highly praised the presentations led by Jo Robinson for its supportive nature and useful insights into managing DIBELS data. They also said that she was inspiring and good at "lending optimism and so much knowledge."

> *The Jo Robinson presentation helped me know how to guide teacher's data collection for DIBELS; I adapted Jo's form so that homeroom teachers look at data and can identify where students need the most growth. (Principal)*

While overall, training was positively received, about half of principals commented that the trainings were too frequent, too repetitive, and/or contained "information overload" that was not always practical (events that were cited included the summer institute, monthly ADE meetings and Open Court presentation by an LA educator). Comments indicated that principals needed more time between trainings to absorb and apply their new knowledge.

> *We had Reading First, Data in a Day, and the entire WestEd professional development agenda. We could only absorb so much. We need to focus on performing well with what we have learned so far. (Principal)*

Coaches

In interviews, coaches overwhelmingly indicated that the training provided to them by the ADE had been very helpful. The monthly meetings were cited by most coaches, with the vast majority of comments being very positive regarding their usefulness. Coaches reported that the meetings had been "invaluable," "fantastic" and that the ADE gave coaches the "necessary tools to do the best possible job."

A strong theme regarding the usefulness of the coaches meetings was that they provided information and strategies that coaches were able to take "right back" to their schools. Similarly, coaches felt empowered by the "train the trainer" aspect of the meetings, expressing that the ability to pass on what they had learned from the ADE to teachers at their school was a "key benefit." These sentiments are exemplified by the following:

> *Training from the ADE has been outstanding because I learned what to bring back. I have been able to replicate the training provided with our teaching staff, which has helped get them on the same page. (Coach)*

Another theme regarding the usefulness of the coaches meetings was that they provided a valuable mechanism for coaches to network and share with other coaches. As the coaching role – as well as the expectations of the grant itself – was new to most, during year one it was perhaps particularly important for coaches to develop relationships with others in the same professional position. As one coach noted, "We hear things that relate to us and we feel that we are not alone."

A small number of coaches had mixed reviews about the training provided through the ADE. A few felt that while the materials and presentations were good, they were repetitive for those who already had a background in reading. Several also noted that the usefulness of the meetings was inconsistent, with some being applicable to their school and others not.

When asked to cite a specific example of a training or strategy that was particularly useful, coaches were extremely positive about trainings provided by Huck Fritterer from WestEd. Almost two thirds of interviewees cited Mr. Fritterer and/or his trainings, most frequently the coaching classroom teachers training or the observing teachers training, and referred to them as "the most valuable" and a "great resource." Many noted that what they appreciated was the practical application of the material and the ease with which they were able to translate what they had learned to the school setting:

> *He gave me something I could take back to my teachers and use almost instantly. (Coach)*

> *Huck is a great trainer, he models the lessons well enough for us to come back to our school and repeat it back to teachers. (Coach)*

In addition, coaches were very pleased with the DIBELS trainings presented by Carrie Hancock from the ADE; approximately one third of interviewees cited these trainings as being the most useful. Coaches appreciated both the sessions on administration as well as analysis and

interpretation. A few indicated that they liked that the training helped them learn how to encourage teachers to "own" their data and use it to guide instruction:

> *It was great, I love being on the cutting edge. Teachers have totally bought in, the more they know, the more eager they become. Now they are anxious to do progress monitoring. (Coach)*

Assessment Coordinators
Almost all assessment coordinators had positive feedback regarding the professional development that they had received through the state; they found it "useful", "excellent" and were "amazed with the quality."

> *The training has been what we needed and what we asked for in addressing the needs of those "in the trenches." (Assessment Coordinator)*

Assessment coordinators noted that the training they had received was particularly critical to the assessment coordinator position, notably the DIBELS sessions; as one commented, "The training is what makes it valid, you can't do it without the training." They appreciated the hands-on activities, which allowed them to learn how to perform this core function of their job.

Assessment coordinators also found it helpful to network and share with other assessment coordinators at the trainings. Particularly for those who work in smaller districts, this was a time of key interaction with others who perform the same "niche" job in their schools:

> *Talking with the other AC's during the meetings about what they are seeing – and realizing that they have some of the same concerns as I do – was very beneficial. It can be a pretty lonely position, so to share information and situations that we came across gave us a bond. (Assessment Coordinator)*

When asked in interviews to cite a specific example of a training or strategy that was particularly useful, assessment coordinators were extremely positive about the DIBELS training provided by Carrie Hancock; in fact, all but two of the interviewed Assessment Coordinators cited this training. Respondents found it "hugely helpful" in their day to day tasks and appreciated both the skill of the presenter and the practicality of the information. Assessment coordinators noted that in the training they learned how to administer DIBELS, enter DIBELS data, use DIBELS benchmarks, and interpret DIBELS graphs:

> *Carrie is personable and her training was a practical, applicable "how-to" on administering the DIBELS. (Assessment Coordinator)*

> *I learned how to use data for individual student progress monitoring, for assignment to intervention groups, and for helping teachers form instructional groups within the classroom. (Assessment Coordinator)*

> *I was able to "get" the DIBELS system right away. (Assessment Coordinator)*

Assessment coordinators also found the Using Assessment to Drive Intervention training with Susan Hall useful for their position. They appreciated that the training helped them learn how to interpret data; likewise, they appreciated that it provided specific interventions for specific skills and grade levels.

> *Susan Hall was great at taking the DIBELS booklet and examining it to get a summary of each student's profile and targeting specific skills. (Assessment Coordinator)*

> *The "Interventions" presentation was very useful to my job. Susan Hall had good presentation skills, knows how to engage an audience, and listens to suggestions. I was able to get out my frustrations and point to some overall weaknesses. (Assessment Coordinator)*

Teachers

Most staff comments on professional development focused on the Summer Academy. Opinions were split between more experienced staff, many of whom felt that the Summer Academy was "a waste of time" without enough "practical examples", and newer staff who felt it was "incredibly helpful" because it provided "a framework" on which to base instruction.

> *The state sponsored Summer Academy came across as a joke for veteran teachers. The consensus was that it did little to further us in the classroom. (Teacher)*

> *I really liked the examples modeled at the Summer Academy. They provided us with real scenarios and talked about how to deal with them, giving lots of concrete ideas. (Teacher)*

Many staff also said that they liked the DIBELS training they received because it helped them "learn to be more data driven" and it helped them design better interventions. The state trainings on instructional grouping and student engagement were also well received by staff because they provided useful teaching strategies to implement in the classroom.

> *The student engagement training was helpful because before we just called on hands and now we have different ways of calling on students. (Teacher)*

District staff

Many of the district representatives interviewed specifically commented on the "great training." Most reported attending some or all of the principals' training and at the Summer Academy, which was said to provide a "good overview." Smaller numbers were reported for a variety of other trainings. For the most part these trainings were rated as very helpful, and "the training has been very good because it gives a similar focus" to schools.

> *There has been lots of training . . . .We are all in the know because of all of it. The state has been very supportive with answering our questions, provided help in visitations, visited the district, let us know where we need to improve. (District coordinator)*

Professional Development Offerings at the School - and LEA - Levels
Beyond the summer academies, most professional development directed at classroom teachers occurred at individual schools or LEAs. In the site visit pre-survey, coaches were asked to list these offerings. In interviews and focus groups, staff members provided feedback about site-specific professional development. This section describes these data.

In the Reading First model, a key aspect of the reading coach role is to provide professional development to teaching staff, often by relaying and disseminating information that they have gleaned from the state or other trainings to staff at their school. Reading coaches also model lessons, observe classrooms and provide teachers with constructive feedback.

Coaches at the 23 visited schools indicated that they provided training sessions on a variety of topics for staff. The most common subject for these presentations was DIBELS; over half of visited coaches noted that they provided professional development to school staff in DIBELS. Other topics included using stations/centers, student engagement, word walls, the "Five Big Ideas in Reading," and sessions on phonemic awareness and/or phonics.

In focus groups, teaching staff indicated that, of all professional development that they received, they were most impressed by the professional development provided at the school-level by the reading coach. In particular, they appreciated the coach's modeling of lesson plans in the classroom and the help the coach provided gathering classroom materials.

> *The coach is good at modeling which I appreciate and another thing that has an impact is that the coach makes materials ready to use and gives them to the teachers. (Teacher)*
>
> *It's nice to get help from the reading coach to put together classroom instruction. (Teachers on a Reading Leadership Team)*
>
> *The coach provided us with structure and materials for centers and word walls. This was great!!! I never had this before – I always gave students seatwork and just had two groups, high and low. Now for the small group time, my groups are better established and others have things to do, they can move among the centers. (Teacher)*

Staff also appreciated the learning opportunities provided at grade-level meetings, which the reading coach often facilitated. Some reported they liked "sharing ideas" and "talking about what they have tried." A few commented that it was important to "reflect" on past professional development sessions during these meetings in order to plan future sessions that met their specific needs.

At some schools, information on reading instruction was passed to staff by the Reading Leadership Team, a team of teachers, the principal in tandem with the reading coach and/or assessment coordinator, or outside consultants. Topics reported by schools included student engagement, literacy centers, observation tools/walkthroughs, and interpreting/using DIBELS data.

Other professional development was provided by representatives from the publishers of a school's chosen core program. By far, the most frequently cited publisher that made presentations at Arizona Reading First schools was Harcourt. Most sessions led by Harcourt provided an overview of the Harcourt Trophies core reading program and its use in the classroom. Other publishers that provided trainings at site visited schools included Houghton Mifflin, Voyager, and Open Court.

Some staff reported that they preferred the professional development offered by their LEA over the training offered by the publishers because this training was more appropriate to their particular school context.

> *District staff development has been fantastic because it has been specific to the school, as opposed to Harcourt and WestEd professional development that has been more general and therefore somewhat less effective. (Reading Leadership Team)*

Potential training topics mentioned by at least one principal were: how to work with ELL students, walkthroughs, differentiated instruction, the coaching role, time management, presentation skills, leadership, grants management, and teaching best practices.

On the other hand, a few principals said that the trainings were "overkill," covering basic information that they already knew or rehashing the same materials each month. Some also felt that the meetings took too much time away from their buildings, either because the material was not worthwhile or because they simply needed time to absorb it and apply it at the school-level.

> *I don't think we have any more time to spend on professional development. We can only absorb so much. We need to focus on performing well with what we have learned so far. (Principal)*

Reading Coaches

In interviews, about half of coaches said that they would like additional training in coaching methods. As many were new to the coaching role, they requested more definition of their role and assistance in implementing it. For most, the role of coach was new to the individual as well as the school. Several noted that they had not received job descriptions until March; at least six months after most took on their coaching role. One coach articulated how this would have been helpful earlier in the year to provide "back-up" and justification of the new coaching role among the teachers at her school:

> *We could have used the job descriptions earlier. Not just for ourselves as coaches, but also to share with the teachers, who thought that we would be reading specialists working with low-level kids and then resented us when we didn't do that. (Coach)*

Likewise, a few voiced their need for the monthly trainings to focus more explicitly on coaching methods and the "how-to's" of their position. For example:

> *We need help with how to do an observation and give feedback. We need this more than we needed our recent introduction to Bloom's Taxonomy. We are not sure if we are*

> *saying the right things to teachers. We need to see it done. We need modeling, just like the kids: I do it, we do it, you do it. (Coach)*

One suggestion was to observe others ("expert coaches") coaching to "see how it's done," rather than only hearing about it in a presentation. A particular area where this could be helpful is in observing how to provide feedback to teachers; one coach expressed the sentiment that, having worked so hard to build rapport with the teachers at their school, they were afraid of being critical and losing their support. Some interviewees suggested that this could partially be addressed by allowing additional time or structures for newer coaches to meet and network with more experienced coaches.

Another suggestion made by coaches was that the ADE allow more time for coaches to talk in small groups and collaborate with colleagues, sharing information about their schools and experiences. Although most coaches were very pleased with the quality and nature of topics discussed at their monthly meetings, a few indicated that they were overwhelmed with the amount of new information. They suggested keeping the trainings simple and/or holding fewer meetings so that they could "focus on implementing what they had learned so far."

Another area in which coaches indicated they would like additional training was interventions. Specifically, coaches were interested in learning more about how to choose an intervention program, how to train staff to use it correctly, and what the intervention program should look like when it is up and running.

Other areas in which a number of coaches noted they would like additional training included: cognitive models of coaching, group facilitation skills, and effective modeling techniques. A few coaches indicated they would like additional training in classroom management models, student engagement, ELL strategies, differentiated instruction/flexible grouping, and integrating the "big picture" of Reading First with what is going on at the classroom-level.

Assessment Coordinators

Assessment coordinators requested more training in various aspects of working with data. The most frequent request was for help with data presentation, including: communicating data to teachers, helping teachers "own" their data, and instruction in how to read the "new" DIBELS graphs. Others were interested in data manipulation (such as transferring demographic data into DIBELS), data entry, and database development/data management.

Another theme that emerged with a few assessment coordinators was that they would appreciate knowing more about the "big picture" of Reading First and overall program goals. Specifically, interest was voiced in learning more about the theory behind Reading First; one assessment coordinator posed the question, *"Why are we doing this?"* In a similar vein, some interviewees indicated that they would appreciate more opportunities to network with other Arizona districts or even other states; this would help them gain a bigger picture of Reading First as well as provide a great opportunity for sharing experiences and lessons learned:

> *It would be great to see other states and/or districts and how they are approaching [the Reading First grant] and any issues that have arisen. (Assessment Coordinator)*

It is interesting to note that interviews revealed something of a split between those who wanted greater involvement with the ADE trainings and those who wanted less. Although interviewers did not specifically ask about this issue, several assessment coordinators and reading coaches voiced their opinions regarding whether assessment coordinators should be invited to more – or all – of the monthly meetings for coaches. On the one hand, increasing assessment coordinator attendance at the current meetings might serve to solidify their involvement in the grant and increase their knowledge of the "big picture" of Arizona Reading First. On the other hand, a few assessment coordinators noted that although the content of the monthly coaches meetings is good, they find that it is not applicable to their role and are not eager to be invited to more such meetings:

> *Some of the topics didn't apply to my job. Discipline problems? I am not in the classroom. They were somewhat interesting and well-done, but not something that I could use. (Assessment Coordinator)*

Teachers

In interviews and focus groups, site visitors also gleaned information on which of the five essential components would be beneficial for teachers to have additional professional development. Vocabulary, comprehension, and fluency were the top three components pointed out as areas in which a need was indicated for additional training. Some teachers also expressed an interest in receiving additional training in classroom management, student engagement, using data and assessment to guide instruction, and working with English language learners.

Another strong concern, noted by principals, coaches, assessment coordinators or teachers at approximately one quarter of schools, were the financial and time costs associated with attending ADE events such as the summer academy and regular monthly meetings. Most felt that while the content was worthwhile, the travel itself was tiring and "burdensome." Schools and districts at further distances from Phoenix were much more likely to feel that the distance required to travel was onerous in terms of time and money, particularly for the monthly meetings:

> *I want the ADE to know that my school spent a lot of grant funds on travel to Phoenix this year; these funds could have gone to support school implementation. (Coach)*

Suggestions included: offering trainings in other regions, spreading the cost of travel across all Reading First schools to achieve a higher degree of equity, and teleconferencing or distance learning options for more distant schools. Another suggestion was to reduce the amount of free/break time during the trainings themselves (e.g. one interviewee cited the breakfast snack, long lunch, and long afternoon as "wasted time") so as to move through the material at a faster pace, which would allow attendees to leave Phoenix earlier and shorten the long day.

Support and Technical Assistance

Communication
In addition to professional development described in the previous section, the ADE supported schools through ongoing communication and technical assistance. In interviews, principals and coaches overwhelmingly praised the ADE for the support it provided them in the first year of the grant. They indicated that the ADE was a valued resource and cited many strengths of their relationship.

Principals, coaches, and some district staff indicated that they felt very supported by the Reading First program staff, whom they referred to as "excellent." They were impressed with their responsiveness and availability, noting the "open atmosphere" which made it "easy to ask questions." Moreover, they felt that these questions were answered in a timely and courteous manner. These sentiments were typified by the following statements:

> *I've been able to find and secure support from them whenever I've needed it. Anything I've wanted to know, I've been able to find on the web or by calling them. (Principal)*
>
> *Whenever I have questions, they answer; they have always been very responsive. They support us 100 percent. They are also great about guiding me to information. (Coach)*
>
> *They have been great! Outstanding. I've never received this type of support before. If I don't get a response immediately, I always get the answer as quickly as possible. (Principal)*
>
> *We have great support and contacts at the state. State Reading First staff responds immediately when we have a question or a concern. I can honestly say that this is a change for them; they are providing "exceptional" customer service on behalf of Reading First. (District staff)*

Another strength noted by principals and coaches in interviews was the ADE's respect for Reading First schools and their staff. This included both the ADE's willingness to accept input from grantees as well as their responsiveness to the individual needs of schools. Likewise, interviewees noted that they appreciated the ADE's ability to balance the structure of the grant and its many requirements, which can feel burdensome, with respect for site-based decision-making and autonomy. One district respondent was particularly impressed with planning at the state level:

> *The state has a good plan in place. It seems that they are taking each step in a calculated, purposeful way. (District coordinator)*

There were very few complaints about technical assistance from the state project staff, but one issue that did come up at several schools was communication. There were two general strands in this area. First, some schools noted a lack of communication between the State and the districts, which in turn resulted in conflicting information and directives for the school. There was an overall sentiment that districts often lacked the "latest" information, perhaps because they were

not "kept in the loop" by the State. For some, this had led to different interpretations of important grant issues – such as budget and Reading First expenditures – between the State and their district. At its highest level, the State/LEA communication issue pointed to a larger conundrum of hierarchy and reporting relationships:

Second, some schools noted that there were communication issues within the ADE itself. Interviewees noted, "You never know who knows what," and "You get different answers depending on who you ask." In turn, some indicated that this results in inconsistent information coming from the ADE and confusion among grantees. Interviewees requested greater clarity, consistency and specificity in communication from the ADE.

Some district representatives foresaw specific needs for the coming year. Many districts reported the need for additional materials; supplemental and intervention materials in particular were widely cited. Districts looked to state project staff for guidance in the selection of these supplemental resources. Additional needs included more in-depth work with data, support for ELL students, looking at the needs of individual children, and the desire for state support tailored to "the unique needs of our campus."

County Reading Specialists

To strengthen their capacity to provide technical assistance to schools, the ADE hired 17 County Reading Specialists (CRS) from across the state. According to the ADE, CRS's should provide support to school leadership in grant implementation, guiding schools in terms of fidelity and the intent of Reading First; they served as "outside eyes" who could come into the school with a bigger picture. They could help schools or districts with specific problems with program implementation. They also served as mentors to school leaders – principals, coaches, and the Reading Leadership Team.

According to the ADE, the priority role for the CRS should be to work with the school leadership, including the principal, coach, and the assessment team. During their visits to schools, they might go into a class with the coach and help the coach learn how to best support a teacher. But the ADE clarified that the CRS should not take the site-based coach's place or assume any site-level responsibilities. Except in the case of an inexperienced coach, the CRS were not to share the coach's role.

During site visits, the reading coach and principal were asked to describe the role of the CRS in their school. Most CRSs provided guidance to coaches and principals on a variety of topics. For example, they helped them create schedules for reading, advised them on observing classrooms, and shared strategies for conducting meetings. Some CRSs attended Reading Leadership Team meetings. Often, CRSs played an important role in supporting reading assessment by assisting with DIBELS training, administration, and data interpretation.

Many CRSs also worked directly with teachers. Some provided trainings on topics such as how to use centers or word walls, small group instruction, and classroom management. Some CRSs attended grade level meetings, conducted classroom walk-throughs, modeled lessons for teachers, and observed teachers and provided feedback. In one school, the CRS facilitated a book club.

In some of the examples above, the CRS filled gaps in areas where coaches could not provide expertise. For example, when CRSs conducted walk- throughs with a coach or principal, they reflected the state's vision of the CRS role because they were working with the grant leaders. It was difficult to determine if the role of CRS overlapped with the role of the coach too much, but data suggested that this was a possibility in a few schools. For example, one CRS reportedly modeled a lesson for each K-3 teacher. This may have been appropriate if the CRS had particular expertise or if the coach was learning how to provide model lessons through multiple examples. However, it may have been a responsibility the coach could have taken on after the CRS modeled in just a few classrooms.

Most principals and coaches noted that the CRSs were "beyond phenomenal" in the diversity of ways that they supported school staff.

> *The CRS is a blessing to our school. I hope she doesn't leave because she is part of our family. She spends time in classrooms and teachers ask for her input. (Principal)*
>
> *The CRS has been of particular help to the reading coach. She worked with staff members and reinforced what they coach had done in training. She provides great moral support to the coach. She meets with me constantly. (Principal)*
>
> *The CRS has been a 'coach for the coach' through this entire process. The coach can freely call her for any questions or training and the CRS responds immediately. (Site Visitor)*

A few of the visited schools reported less positive experiences with their CRS. In some cases, the school felt the CRS could not provide enough services because she was "spread too thin" or only visited the school once a month. Some school staff said their CRS was not familiar with their core program, which limited the amount of technical assistance she could provide. In addition, a few interviewees felt the CRS offered too much praise and too little constructive criticism, or did not have up-to-date or accurate information from the state. Finally, a few interviewees reported that the CRS did not bring new or useful knowledge or services to the school:

> *The CRS has offered a few services like helping administer the DIBELS once and attending some meetings. But we usually don't go to her with questions because we can go to the principal or district. (Coach)*
>
> *The CRS does not have the expertise about reading, or accurate information from the state. (LEA representative)*

County Reading Specialist Experience

Introduction

We have discussed what others have thought of the role of CRS, here is what they say about themselves. This section reports the findings derived from both the short survey and the focus group held with county reading specialists on April 23, 2004. The information from both data sources is combined in this section to present an overall description of the attitudes and opinions of the county reading specialists.

In general, the entire session, and in particular the focus group, may be characterized as cooperative and energetic. Everyone in the room actively engaged in the group discussion, although some more than others. In fact, all the members of the group spoke at least once during the discussion; and no one person dominated the discussion. To some extent, the opinions expressed in the focus group varied on the basis of how the districts, coaches, and teachers cooperated with the reading specialists as they performed their role. Seventeen specialists were present.

The purpose of the focus group was to examine the following: (1) the role and responsibilities of the county reading specialists; (2) the perceived changes and successes in the Reading First-schools in the past year and the role they played in the changes and successes; (3) the challenges of the role; and (4) the degree to which the training they received prepared them for the role as reading specialists.

Before the focus group discussion, the reading specialists were administered a short seven-item survey to obtain information on questions similar to those addressed in the focus group. The purpose here was to generate responses from individuals to see whether they differed from the group responses in order to isolate the group's influence as well as to provide a quantitative framework for presenting the discussion. Other items unique to the specialists were also included in the survey. There were 14 out of 16 of the specialists whom completed the survey, as two were absent from the group.

According to the ADE, the role of the reading specialist is to provide content expertise in reading instruction as needed to assist schools in implementing the program and monitoring their progress. The reading specialists provide support and assist teachers, schools and districts in the implementation of Reading First, focusing on enhancing the level of expertise of K-3 reading teachers through the application of evidence-based strategies in reading instruction. The responsibilities of the reading specialists are to: (a) serve as the local representative of ADE, (b) engage in professional development sponsored by ADE, (c) provide technical assistance to the districts and schools in his/her region, (d) provide expertise in SBRR and its implications for classroom instructional practice, (e) and provide ongoing support and assistance to schools in addressing reading achievement in his/her region.

In this role, the county reading specialists are pivotal in addressing several of the major purposes of Reading First and their assessments inform four of the evaluation questions:

1. Professional Development: Knowledge Transfer to Teachers
2. Transfer of Knowledge in the Classroom

3. Capacity Building: Support System
4. Capacity Building: Leadership Development

The survey and focus group facilitator was aided by two recorders, one to take notes of the proceedings and the other to record on a flip chart the topics and issues generated by the group for review. The focus group session lasted for approximately 40-45 minutes. For subsequent analysis, the evaluator also audiotaped the group session. The County Reading Specialist survey and the Focus Group protocol are located in Appendix J.

Role and Responsibilities
The majority of the county reading specialists (8 of 14) said they served 3 or 4 schools; three were assigned to five schools, and one each served six, seven or ten schools. Most served schools in two or more districts. The schools served were made up of both Reading First Schools and non-Reading First Schools. One specialist was assigned 7 Reading First Schools, which was the maximum.

Table 7.11 shows the average number of days per month that county reading specialists spent at each school. The large majority of the reading specialists (10 of 13 respondents) indicated that they spent 3-4 days at each school in the average month. Two said they spent 1-2 days, and a one respondent spent 5-6 days at each school in an average month. The two specialists who spent the 1-2 days also served the highest number of schools.

Table 7.11
Average Days Per Month at Each School

	Number of Average Days
1-2 days at each school	2
3-4 days at each school	10
5-6 days at each school	1

When asked whether they were able to spend the time doing what they needed to do in their role as reading specialist, the large majority (9 of 13) indicated that they spent their time "mostly as needed," and two said they spent "very much" of their time as needed. The other two spent their time "somewhat as needed." The results are summarized in Table 7.12.

Table 7.12
How Time Was Spent

	Number of Responses
Very Much as Needed	2
Mostly as Needed	9
Somewhat as Needed	2

The county reading specialists discussed that they provided the content expertise and technical assistance to the districts and schools as their role called for and were effective in doing so. For the most part Specialists spent their time with the reading coach who is their main contact. Since they do not have a defined role for every school this can vary and is ultimately up to the Principal. Evidence of this is captured by a specialist's own words, "It's amazing how far coaches and schools have come. I feel we have been a catalyst in this process." In fact, the reading specialists, as a group, indicated that they had no regrets taking the position, for the benefits they received out-weighed any of the problems that may be linked to the role. This was due mainly to the benefit of seeing the positive results from the specialists' involvement in the schools.

Much of their time away from the school was spent traveling, attending meetings, doing prep work for in-school workshops, or simply following up a question or request from a coach. With one of their key functions being progress monitoring, it then required the specialist to perform daily logs (who, where, what, when) and then report to ADE twice a month per school.

Successes

As reported in Table 7.13, in the past academic year, six of the reading specialists said their Reading First schools had "very much" or "mostly" achieved the gains and successes expected of them. However, the majority (7), albeit only a majority by one, indicated that the schools made only moderate (i.e., "somewhat as needed") gains and successes in the past academic year.

Table 7.13
Schools Achieve Gains and Successes

	Number of Responses
Very Much	1
Mostly	5
Somewhat	7

The reading specialists were also asked how effective they believed they were in contributing to the specific gains made by the schools under their charge, as shown in Table 7.14. Four of the reading specialists said they "very much" had a role in the gains made by the schools. Seven of the specialists said they had "somewhat" of a role in the gains made by the schools in the past academic year and 3 said they "somewhat contributed" to the gains made by the schools.

Table 7.14
County Reading Specialists Contributed to Schools Gains

	Number of Responses
Very Much Contributed	4
Contributed	6
Somewhat Contributed	3

CRS generally thought they were effective in many aspects of the RF implementation. By creating expectations, the specialists were able to see the growth of teachers by moving them away from old habits of teaching on instincts and instead adapting their lessons based on the data. In one particular school this produced far fewer special education referrals. The ability of the specialists to shift attitudes, not only of the teachers, but of the principals as well, was essential to RF implementation. As one participant noted, "At first Principals and Teachers balked at the high expectations, especially for 1st graders, but by Christmas after seeing the results they were on board." Another example of this was "We (CRS) are now perceived as being helpful and nonjudgmental."

The following represent the major points the reading specialists identified as their successes.

- County Reading Specialists were able to problem solve together;
- More students are learning and less are left behind because teachers are moving away from old habits of teaching on instincts and instead wanting to learn from data and then seeing the students' growth;
- CRS positive change agent: Helpful in school as a support system and means of communication with ADE.

On the survey, the specialists were given a list of challenges related to the role of the reading specialist and asked to check the 3 greatest challenges they faced this year. The "lack of teacher buy-in to the Reading First goals" ranked first on the list of seven items checked by the specialists. "Lack of time to provide all the services requested or needed ranked 2nd, followed by the "location of or convenient access to the schools," in third place. The results are provided in Table 7.15.

The specialists who wrote-in comments listed the following challenges: (1) the lack of communication within the district regarding expectations and requirements of the Reading First program; (2) questions about the validity of the implementation checklist after the decision was made to share it with the schools; (3) too many schools to be effective in their role as reading specialist and (4) scheduling of grade level meetings had to be insisted upon for next year.

Table 7.15
Three Biggest Challenges Faced

Challenges Faced	Number
Location of or convenient access to the school	6
Difficulty matching your schedule to the school's	3
Lack of time to provide all the services requested or needed	7
Lack of teacher buy-in to Reading First goals	9
Difficulties with one or more reading coaches	2
Difficulties with one or more assessment specialists	1
Difficulties with one or more principals	4
Difficulties with one or more school districts	1
Other	4

For the most part, the problems discussed with fulfilling their roles were relatively minor and mainly consisted of geographical issues that faced specialists in rural areas. The reading specialists had concerns regarding the travel distance in the rural regions. In fact, the distance in miles they traveled to the schools affected their morale, i.e., the more time they spent traveling, the less time they had in the classroom. Hence, the time spent traveling in the district created not only emotional problems (being away from home more than anticipated), but also financial hardships of the longer than expected reimbursement process.

The specialists also had problems with the teachers as they adopted the core programs in their classrooms: the use of the core; the static progress due to failure to adapt the core to the needs of the classroom; adhering to the core like robots with too little creativity; and keeping teachers faithful to the core but allowing for change. This problem may be due to the strict schedule, which was sometimes implemented district wide, that tells teachers where they should be, so they are likely to follow the teacher's manual closely. In other cases, the teachers were viewed as resisting and hanging on by questioning leadership every step of the way. This was clearly heard from one CRS who stated: "Certain teachers are not progressing, they just follow the core without adapting to the needs of the classroom. They become like robots with too little creativity. This comes from a strict schedule, sometimes-implemented district wide that tells teachers where they should be. Challenge is keeping them faithful to the core, but allowing for change." A factor in this was no doubt the learning curve involved for teachers in it being the first year the core was implemented.

There were two main surprises that CRS encountered as result of the grant. One was that some of them were relocated to Maricopa County. For example one respondent stated, "I was under the impression from ADE that I would get schools within my own county. Much of my time is spent traveling and not enough at my schools." The second surprise they faced was the amount of time necessary to build relationships among the key players. " The Principal has not looked at me as an equal, they would check up on me. They need to know to trust us and not go over our head."

The following are the major points the reading specialists identified as their challenges.

- Lack of Communication, information about expectations from schools (principals, teacher) from district;
- Resistance to Program in some areas;
- Staying too faithful to core – teachers follow it point by point & fail to use creativity to improve lessons.

Reading specialists identified two major challenges for next year. First, they believed they were spread too thin by the amount of paper work and meetings they had to attend to fully meet the needs of the schools and the expectations of ADE. As one CRS noted: "How can we tighten the reins when we're in the schools so seldom?" The second challenge was the amount of traveling which caused them to be away from their families and a financial strain of fronting the costs involved.

Training/preparation
The last area covered by the focus group and the short survey asked the specialists how well they felt they were trained or prepared to provide expert services to the schools. Table 7.16 shows that slightly less than half (6) said they were "well" or "very well trained/prepared" to provide expert service to the schools. Seven were less forceful in their response and indicated that they were "somewhat well trained" for the reading specialist role.

Table 7.16
How Well Trained/Prepared

	Number of Responses
Very Well Trained/Prepared	5
Well Trained/Prepared	1
Somewhat Well Trained/Prepared	7

Their comments were representative of the major points discussed by the reading specialists as to their training.

- The training was very helpful, even if there was not enough of it to meet all of the challenges (building trusts) that were counted on;
- Not trained to do some of the things;
- We felt less than adequate; we just didn't have the training.

Conclusion
Although the focus group discussion did not cover or capture all the issues that the county reading specialists faced in supporting the districts and schools, the comments suggested that there are numerous areas in which positive changes and growth had occurred. Even so, it is obvious that there are still problematic areas that need to be addressed further. Since the objective of the focus group was to obtain information from the reading specialists regarding their role in assisting districts and schools implement the Reading First program, what they had to say should be helpful for improving the program.

The focus group and survey results were fairly consistent regarding the gains and successes made by the schools in the districts; the extent to which they had a role in the gains; and whether they were adequately trained or prepared to effectively carry out their role and responsibilities. In fact, the survey results compliment and add to the focus group results. Although the results did not differ markedly between the focus group and survey responses, the latter provided the views of the specialists apart from the group, so it was helpful in that regard. Moreover, the survey generated important information on the number of schools the reading specialists served, whether the schools served were located in the same or in multiple districts, and the number of days they spent at each school in an average month.

In both the focus group and survey, the reading specialists qualified their responses as to whether they were "well trained/prepared" in the role of reading specialist to provide expert services to the schools. While they highly praised the training they did receive, they still believed that more

training was needed, indeed more than half felt only "somewhat well trained" for their work during the past year. However, the results leave little doubt as to the perceived contributions of the reading specialists to the gains made by the schools over the past academic year.

In conclusion, the reading specialists' own assessments of the role they played indicated that they believed they were doing what they were assigned to do and needed to do as county reading specialists. They believed that for next year they face challenges with travel and implementation of the core curriculum. Overall, the reading specialists indicated that with additional preparation to come (summer 2004), they will be better prepared for the challenges of the next academic year. Based on their responses in the focus group and in the survey, county reading specialists are in the mindset of meeting the upcoming challenges.

LEADERSHIP AND SCHOOL-LEVEL STRUCTURES

Highlights

The Role of LEAs

- County Reading Specialists reported a variation in the level and type of support districts provided to Reading First schools. The most common role for LEAs to take during this first year of implementation was grant administration and management. Some LEAs also provided technical assistance and material support. Those that provided technical assistance reported that this was more time-consuming than anticipated.

- Representatives from most LEAs indicated that the Reading First project had an impact on many or all the elementary schools in their district, not just the Reading First schools. Some LEAs had provided reading training for non-Reading First schools, or encouraged schools to use the DIBELS assessment, while others had made scientifically-based reading materials more broadly available.

- Most of the schools visited expressed appreciation for the support of the grant from their LEAs, support that came in the form of meetings, sharing of knowledge, or arrangement of substitutes to facilitate training attendance.

- About a third of schools described some problem or frustrations in their work with LEAs on Reading First, sometimes because of poor communication or a lack of understanding about what the LEA role should look like.

Leadership and Key Roles

- Staff at most schools described their principals as "very involved" in the implementation of the Reading First grant at their school. Most principals met regularly with teachers and the coach, and participated in reviewing and sharing assessment data. Almost all principals conducted classroom observations and attended grade-level meetings, although these were the lowest-rated items on the implementation checklist.

- At all but two schools, the working relationship between the coach and principal was positive and in many cases was a very close collaboration.

- Coaches were most likely to list observing classrooms, mentoring teachers, conducting demonstration lessons or other professional development, and assessing students or managing assessment data as the tasks which took most of their time.

- While supporting teachers and working on the assessment team were primary responsibilities, most coaches were fulfilling the other roles specified on the implementation checklist and in interviews they provided a long list of other responsibilities. For nearly half of the coaches, a shortage of time to complete all the tasks was a big enough concern to mention in the interviews.

- On surveys, coaches reported that they were confident in their role as coach and that this confidence increased as the year progressed. Despite their comfort with their roles, another serious challenge for a substantial number of coaches was resistance from teachers. This ranged from building trust with teachers unaccustomed to classroom observations to struggling with openly hostile teachers who did not buy into the Reading First program.

- Nineteen of the 23 schools visited, or 83 percent, had an assessment coordinator at the school, although ten of those positions were part-time. The main task of assessment coordinators was to coordinate the administration, data entry, and interpretation of the DIBELS benchmark assessment.

- Responsibility for a range of other tasks, including training teachers about assessments, working closely with the Reading Leadership Team, and managing progress monitoring, meant that assessment coordinators listed time constraints as their number one challenge.

- Most assessment coordinators were excited about the growing interest in and reliance upon assessment data at their schools, although a few reported that low assessment scores made them personally unpopular at their schools.

- Reading Leadership Teams (RLTs) at the 23 schools visited consisted of the principal, coach, and in schools that had the position, the assessment coordinator and teachers, usually one per grade. Special education and Title I teachers sometimes participated but rarely were paraprofessionals or parents part of the Reading Leadership Team.

- A few schools had highly functional RLTs that actively reviewed data together, made decisions and worked to ensure program fidelity. Many schools had teams that served primarily as information sharing structures; their role was very much still evolving. On the other end of the spectrum, a few schools had teams that met rarely or not at all.

Communication, Collaboration, and Support for Reading First

- Support for Reading First was fairly high at the start of the grant; 71.6 percent of matched survey respondents indicated strong support and 26.2 percent indicated moderate support to a survey statement about supporting instructional changes that occurred under Reading First. However, support declined slightly over the year (to 63.7 percent and 31.8 percent respectively). Data revealed a larger decline in support among third-grade teachers compared to other grades.

- These levels of support for Reading First were fairly evenly spread across schools. A handful of schools reported very positive experiences with high teacher buy-in; conversely, a handful of schools had relatively poor experiences, with teachers displaying "serious" resistance.

- Across the board, schools reported improvements in communication and collaboration over the past year; however, the degree of improvement varied greatly from school to school.

- In general, those schools with higher levels of buy-in for Reading First at the beginning of the year indicated greater successes in increasing communication and collaboration, compared to schools that had lower levels of initial buy-in.

- One of the most commonly cited challenges across the evaluation was teacher resistance; developing ways to draw teachers into the program – and identifying and working with the schools that struggle most in this area – may be an area in which support from the ADE coming year could be beneficial.

Local Education Agencies (Districts)

The 63 Reading First schools are located in 27 Local Education Agencies (LEA) across Arizona. The ADE envisioned the LEA as a source of support for Reading First schools and principals in their implementation of the grant. According to the ADE, the LEA should provide clear support to schools' Reading First goals and plans. In addition to assisting with financial elements of the grant, LEAs were encouraged to provide support through cooperation and flexibility, such as extending the window for purchasing intervention materials to allow for thoughtful decision making.

County Reading Specialists (CRSs) reported that LEAs varied in providing support to school, as measured by five items in the implementation checklist (Table 8.1). Specifically, CRSs reported that about half of the schools were supported by LEAs that had fully implemented district supports for school-level RF activities and had provided technical assistance as needed; the rest showed "some" or "no/scant" evidence. Ratings for at least a dozen schools moved from "no/scant" to "some" or "full" implementation of these items between February and May.

Additionally, CRSs found that Reading First activities in about one in five schools were not monitored by their LEA and a similar number of schools were not located in districts that provided Reading First activities to non-Reading First schools.

Table 8.1
Local Education Agencies (Implementation Checklist)

	May 2004 Percent (n)			Change from February*
	No/ Scant	Some	Full	
The district supports school-level RF activities with adequate resources.	6.5 (4)	41.9 (26)	51.6 (32)	+ 14 schools
The district provides technical assistance as needed.	11.3 (7)	33.9 (21)	54.8 (34)	+ 12
The district leadership is involved in the coordination of activities among RF schools.	14.5 (9)	45.2 (28)	40.3 (25)	+ 8
The district monitors RF activities and intervenes when necessary.	21.0 (13)	40.3 (25)	38.7 (24)	+ 10
The district leadership coordinates district-wide RF activities that include Arizona READS schools.	19.7 (12)	45.9 (28)	34.4 (21)	+ 17

*Number of schools that increased from "no" or "scant" evidence to "some" evidence or "fully implemented" from February to May. See text for discussion of limitations.

Role of LEAs

District representatives were asked to describe their role in the Reading First project. As might be expected, the most commonly reported district role was grant administration and management. Interviewees reported that their involvement started in the early stages of grant application including grant writing oversight, budget development, and clarification of state expectations.

While a few of those interviewed indicated that administrative responsibilities were the main focus of their activities, district staff commonly reported that they supported schools in a much broader way:

> *Our role is to support the Reading First project 100 percent with technical assistance, encouragement, providing materials, personnel, and all the supplies they need to be successful. To be a support system for them, to remove obstacles instead of creating them.*

Several district staff detailed the many ways they had assisted schools, including hiring district staff to support the program, helping schools fit Reading First in with other initiatives, and organizing district-level meetings. Several interviewees noted that these roles required a lot of time and effort from district staff (often more than originally expected).

> *Besides support to administer the grant, we monitor, purchase, support training, and keep principals apprised on expectations. The district has bi-weekly meetings of coaches to solve problems in all of our Reading First schools. The district initiated the grant and helped write it. We have also hired two district reading coaches that support Reading First.*

> *It's almost a full time job helping only one school. The district should have been aware to have a full time Reading First director . . . It is a lot more work than what we anticipated.*

Many, although not all, interviewed district staff indicated that the type of support they provided schools under Reading First was substantially different from the type of support they provided schools in the past. One commonly reported example was moving from administration toward areas more direct involvement with instruction and professional development at the schools.

> *In the past we coordinated textbook adoptions, ordering, and related professional development, and then the principals pretty much took over. This time, we are fully involved in all aspects of the grant implementation in an ongoing way.*

> *It has been way different. We [the district] hold meetings every month to go through needs, reports, etc., planning what teachers will do, discussing what they have done. It has really focused the reading project so much more. It helps the whole district to focus.*

In most interviewed LEAs there was evidence that the impact of Reading First was reaching beyond those schools that were part of the grant. CRSs also found evidence of "some" or "full" implementation of this item in approximately 80 percent of schools. However, districts were at different points in spreading out the components of Reading First; several shared trainings (e.g., DIBELS) with other schools, others had purchased materials for the non-Reading First schools, as well as materials for grade levels beyond the grant within Reading First schools, and a few planned to begin similar steps next year. One reported that a Reading First summer school would include non-Reading First schools.

> *They are already using DIBLES at the other two [non Reading-First] elementary schools, we trained them…In addition, the core program was purchased for kindergarten and first grade, next year we plan to purchase it for second to sixth grades.*

> *Because of Reading First, we did a reading adoption for the first time in nine years… We've gone, with our adoption, from three Reading First schools to 25 district schools… We are using Reading First to begin to build a district-wide system of support for scientifically-based reading research and our new adoption.*

Looking ahead to the second year of implementation, a few LEAs planned to allocate additional staffing to the Reading First work, such as hiring an intervention specialist for each school. Several districts reported that personnel concerns were a challenge, as it was not easy to replace unsatisfactory teachers. Some faced the challenge of hiring "strong people" for next year to replace those in leadership roles who had resigned.

Schools' Perceptions of District Support

Overall, the majority of principals, coaches, and assessment coordinators in visited schools were positive about the support provided by the district. One of the most commonly cited supports was "bringing people together", mentioned by both principals and coaches. In one LEA, a respondent reported, the district called a meeting at the beginning of the grant to "set the tone" for Reading First and lay out the expectations. In another LEA, all Reading First principals were brought together monthly for meetings, and a third LEA arranged weekly meetings with the two district reading coaches and all school reading coaches.

> *The school district has invested a lot of money in training for teachers, plus they bring the coaches in the county together once a month, which is great because we face similar challenges and can help each other.* (*Coach*)

In relation to the grant, some interviewees described district staff as "very approachable and responsive," and a "great support and source of knowledge." They reported that the resources provided by the district were particularly helpful, including arranging for substitute teachers so that staff could attend workshops. Assisting schools with assessment, organizing trainings, and leading meetings, were also cited by school staff as helpful roles the district had played.

> *They take the cumbersome things off our plate when they can, like compiling the information for the pre-survey, making things more user-friendly. They are always willing. We are not in this alone. (Principal)*

While the majority of those interviewed felt the LEA had played a positive role in Reading First, about one-third of interviewees indicated some dissatisfaction. Concerns included uncertainty about the district's role, communication problems, and a difficult beginning with little district support. There were indications later in the year of improvement according to two schools with early difficulties.

> *It's been getting better. The district feels overwhelmed and a bit frustrated. They are open and receptive to the responsibility, but lack the time to follow through. (Coach)*

> *Support is good, it could be better in communication about the grant; we're going day by day. Everyone has been learning so it's hard to say, but I would like better communication on things like budget, and I feel like there should have been a plan. (Coach)*

A small number of interviewed school staff reported deeper problems with lack of organization and preparation:

> *The district needs to streamline the process so everything is not all over the place.*

> *We've had minimal support. I tried to get the district representative to come out to the school, but she was too busy and didn't come. You know, we needed her to come out and have HER say to the staff that the district supports us and that they know we are working hard with Reading First. (Principal)*

At times there were discrepancies between the district's description of support, and the view of school respondents, or between respondents from the same school. For example, one district representative offered a lengthy list of ways the district supported Reading First, while a principal from that district reported that, beyond budgets and ordering materials, the district role was unspecified.

School Leadership

The Arizona Reading First project provided training and technical assistance to build leadership in principals, coaches, assessment coordinators, and Reading Leadership Teams. Each of these is described separately.

Principals
Principals played a key role to play in the implementation of Reading First at the school level. It was they who had the position and authority to designate Reading First as a top priority in each school and to set the tone for successful grant implementation. According to the state project director, the principal should be involved in the day-to-day work of the grant. Further, as an instructional leader, the principal should be visible in the classrooms, observing and providing feedback; this entails being able to recognize appropriate instruction. In addition, the principal should provide a consistent message that supports the program, work closely with the reading coach, and be able to lead a collaborative team in implementing schoolwide reading goals and objectives. Information about the particular responsibilities, implementation, and leadership abilities of principals was gathered with the implementation checklist, and site visit interviews.

The majority of CRSs reported that they saw "some" or "full" evidence of most items related to principal leadership (see Table 8.2). These included: attending RLT meetings, reviewing and sharing data, ensuring that the 90 minute block was uninterrupted, planning professional development in reading, and scheduling grade level meetings on a regular basis. As shown in the far right column of the table, these roles solidified in many schools between February and May of Year 1. For example, in May CRSs reported that an additional 20 principals showed at least "some" evidence that they led staff in analyzing data. In interviews, principals added that their responsibilities included attending trainings, working with the coach, and monitoring the Reading First budget.

Table 8.2
Principal Roles (Implementation Checklist)

	May 2004 Percent (n)			Change from February*
	No/ Scant	Some	Full	
The principal provides a master schedule that protects a minimum of 90 minutes for reading instruction.	-	12.7 (8)	87.3 (55)	+ 3 schools
The principal makes decisions necessary to ensure the effective implementation of Reading First goals.	-	31.7 (20)	68.3 (43)	+ 9
The principal ensures identified teachers receive the assistance/intervention they need in instructional practice.	1.6 (1)	39.7 (25)	58.7 (37)	+13
The principal has established a Reading Leadership Team with representative teachers from all grades K-3, including specialists (SPED, ELL).	4.8 (3)	31.7 (20)	63.5 (40)	+18
The principal schedules grade level collaboration meetings regularly.	6.3 (4)	20.6 (13)	73.0 (46)	+ 7
The principal leads the staff in analyzing assessment data to design and monitor instruction.	11.1 (7)	57.1 (36)	31.7 (20)	+20
The principal chairs and provides direction at the RLT meetings.	12.7 (8)	27.0 (17)	60.3 (38)	+15
The principal documents evidence of the implementation of the core reading programs.	14.3 (9)	38.1 (24)	47.6 (30)	+10

*Number of schools that increased from "no" or "scant evidence" to "some evidence" or "fully implemented" from February to May. See text for discussion of limitations.

The items in the implementation checklist that were rated lowest pertained to walk-thrus and attending grade-level meetings (Table 8.3). According to CRSs, although most principals observed classrooms, about one-fourth of schools still showed "no/scant" evidence that the principal conducted *daily* walk-thrus and one-fifth of principals did not attend grade-level meetings regularly. This pattern also emerged from interview data from the 23 visited schools; many principals said they were able to be in the classrooms every day. In a small number of schools, respondents during qualitative interviews noted that the principal was somewhat removed from daily implementation activities. (In these schools, principals were still considered by coaches and staff to be "supportive").

Table 8.3
Principal Observations and Attendance at Grade-level Meetings
(Implementation Checklist)

	May 2004 Percent (n)			Change from February*
	No/ Scant	Some	Full	
The principal observes reading instruction in each K-3 classroom to ensure research-based instruction is sustained.	6.5 (4)	37.1 (23)	56.5 (35)	+9
The principal provides constructive feedback to teachers based on observations and walk-thrus.	19.0 (12)	31.7 (20)	49.2 (31)	+ 11
The principal attends grade level meetings regularly.	20.6 (13)	39.7 (25)	39.7 (25)	+12
The principal conducts daily "walk-thrus" during reading instruction using a research-based reading observation checklist.	25.4 (16)	57.1 (36)	17.5 (11)	+16

*Number of schools that increased from "no" or "scant evidence" to "some evidence" or "fully implemented" from February to May. See text for discussion of limitations.

Site visitors found that the majority of principals were active in day-to-day implementation and monitoring of the grant. Several principals provided in-depth information on their participation in various aspects of the program, such as describing walk-thrus in detail. Many coaches described principals' level of activity as "very involved at every level" or "very committed," sometimes in glowing terms. The coaches below described principals who support grant implementation:

> *Our principal is the heart of the whole school. She's the one with the vision to carry out this project. Teachers know what she wants and they work hard towards it, she is extremely positive with the teachers and in the classroom daily, she works very hard to make sure this will be a positive experience for everyone. She oversees the whole thing, she is the one with a global view. (Coach)*

> *We do walk-abouts together and she makes sure the teachers are implementing the core. She's in the classrooms more than before, she attends literacy team meetings, makes sure that everything runs smoothly and keeps paperwork up to date, and also keeps in touch and discusses her observations with me. She makes sure the teachers are actually participating and willing to do the core program. (Coach)*

> *She is the driving force behind the grant, maintains high expectations, and encourages people to step up. (Coach, with agreement from Assessment Coordinator)*

> *The principal drives the whole program of pushing for reform. His leadership has been essential...Without him we would not be successful. He put the RLT together, which has been very productive. Now everybody has a buy-in to the program. (Coach)*

Some principals reported that fulfilling all the varied aspects of their role had meant a steep learning curve:

> *I spend a lot of time learning how to do this job! Sometimes I feel like the grant is all I think about and dream about! (Principal)*

Site visits provided feedback that there were concerns in a few schools regarding the principal's role. These included the principal's focus on management with little classroom presence, lack of understanding of Reading First, or conflict between the principal and staff.

> *This year's focus has been more on the management of the grant in order to get staff together. Next year I hope to be in the classroom more, using observation check lists and giving teachers feedback. (Principal)*

> *The principal has time restraints because of other duties. (Teachers on the Reading Leadership Team)*

> *She is very knowledgeable…but her knowledge does not align with Reading First. She has evaluated Reading First in each classroom according to her own agenda. She is very avid about scores. She likes to analyze and tell everyone what they need to do. (Coach)*

> *I have been trying not to take things personally, but it's been hard… when the principal makes the teachers cry. (Coach)*

Reading Coaches

Reading coaches played an important role in creating the kind of school- and classroom-level change envisioned by Reading First. While their position involved some leadership and certainly some collaboration with the principal, their primary responsibility was to support and mentor teachers as they adopted research-based instructional practices and materials in their reading classrooms. To this end, coaches were informed by the ADE that about 80 percent of their time should be spent in the classroom and/or working directly with teachers.

In addition, coaches were to play a leadership role on the RLT, help to plan the logistics to establish a schoolwide intervention system, and support the implementation of the school's reading plan. In collaboration with the principal and assessment team, coaches were to help analyze assessment data and help teachers make instructional and grouping decisions using data.

Information about the tasks that coaches participated in was gathered with both the implementation checklist and interviews conducted during site visits. In addition information on prior training, confidence in the coaching role, and challenges faced by coaches was gathered with the implementation checklist, survey, and site visit interviews.

According to CRSs, the majority of reading coaches had assumed the roles and responsibilities measured by the implementation checklist by the first administration of the checklist in February. By May, all or almost all schools showed at least "some" evidence that most coaches were: assisting with assessment; documenting their assistance; assisting with monitoring the use of materials in cooperation with the principal; and assisting with training for the core reading program.

Table 8.4
Reading Coach (Implementation Checklist)

	May 2004 Percent (n)			Change from February*
A job description for the Reading Coach has been established.	98.4 -yes- (62)			+ 2 schools
	No/ Scant	**Some**	**Full**	
A site-based coach has been hired to work with and support K-3 teachers.	-	3.2 (2)	96.8 (61)	+ 2 schools
The coach assists the assessment teams in administering, scoring, recording, sharing, analyzing and interpreting student data.	-	6.3 (4)	93.7 (59)	+ 3
The coach assists the principal in monitoring the ongoing use of research-based practices with approved materials.	4.8 (3)	23.8 (15)	71.4 (45)	+ 5
The reading coach documents the assistance he/she is providing.	3.2 (2)	31.7 (20)	65.1 (41)	+ 6
The coach assists in the identification and implementation of interventions.	9.5 (6)	27.0 (17)	63.5 (40)	+ 9
The coach assists teachers with adjustments to instruction based on data.	9.5 (6)	41.3 (26)	49.2 (31)	+ 10
The coach assists in the ongoing training of the core reading program.	12.7 (8)	31.7 (20)	55.6 (35)	+ 1
80% of the coach's time is spent coaching K-3 teachers in variety of activities.	12.7 (8)	57.1 (36)	30.2 (19)	+ 9
The coach creates and maintains a schedule for coaching teachers.	15.9 (10)	39.7 (25)	44.4 (28)	+ 10

*Number of schools that increased from "no" or "scant evidence" to "some evidence" or "fully implemented" from February to May. See text for discussion of limitations.

Despite fairly high ratings on most items, CRSs indicated that not all coaches were able to spend 80 percent of their time coaching teachers. This was true despite the fact that interviewed coaches identified observing classrooms, mentoring or providing feedback to teachers, delivering professional development to groups (teachers, tutors, etc.), conducting classroom demonstrations, and assessing students and/or managing assessment data as the activities they spent the most time doing.

In addition to those primary tasks, however, coaches described a vast range of additional responsibilities, including the following (those higher in the list were mentioned by more coaches; those near the bottom were mentioned by only one or two coaches):

- Meeting with the principal or RLT, planning implementation
- Obtaining materials for teachers
- Being trained themselves, or learning information they needed
- Placing students in appropriate groups
- Monitoring interventions
- Completing paperwork or keeping records
- Providing and/or monitoring interventions to students
- Scheduling (trainings, County Reading Specialist visits, interventions)
- Trouble-shooting
- Assessing students in intermediate grades

Given the vast array of responsibilities and the time it took to learn the tasks and carry them out, it was not surprising that many coaches noted the long hours that they worked during the first year of implementation and the inability to provide coaching 80 percent of the time. A few coaches even described themselves, or were described by principals, as "assistant administrators" with a focus on reading.

Prior Training and Comfort with Coaching Role: Overall, coaches reported on surveys that they were confident in their role and that this confidence increased as the year progressed. Specifically, they were comfortable at both pre-test and follow-up helping other teachers learn to tailor reading instruction to different levels. (Table 8.5). There was a large increase over Year 1 of the percentage of coaches who said it was "very true" that they had been formally trained to observe teachers and had experience doing so. (Most respondents who did not select "very true" selected "somewhat true" to all items in the table.)

Table 8.5
Coaches' Preparation for their Role (Survey Responses)

	Percent responding "Very True"*	
	Pre-test	**Follow-up**
I am comfortable helping other teachers learn to tailor reading instruction to multiple different levels within the classroom.	70.7%	69.5%
I have been formally trained to work with adult learners.	35.0%	47.5%
I have been formally trained to observe teachers and provide constructive feedback.	30.0%	72.9%
I have experience observing teachers and providing constructive feedback.	48.8%	71.2%

***Responses from 59 coaches with matched pre- and follow-up surveys. Actual "n" for each item may range slightly due to missing cases.**

The survey also assessed the coaches' perception of their role and support for Reading First. Although many coaches lacked formal training in some of the tasks they faced in the school year, all coaches reported that they supported the approach promoted under Reading First (see Table 8.6). The survey also found that coaches were less sure about what Reading First would require in the upcoming year at the end of the first year of implementation. It may be that their responsibilities were quite different from their expectations at the beginning of Year 1, decreasing their belief that they "know what is coming" in the future.

Table 8.6
Coaches' Perception of their Role and Support for Reading First
(Survey Responses)

	Percent Responding "Very" or "Somewhat" True*	
	Pre-test	**Follow-up**
I strongly support the instructional and other changes that will occur under the Reading First grant.	100%	100%
I participated in the conceptualization and/or writing of our Reading First grant proposal.	30.5%	30.5%
I have a good sense of what my role in Reading First will require of me in the coming year.	92.7%	59.3%

*Responses from 59 coaches with matched pre- and follow-up surveys. Actual "n" for each item may range slightly due to missing cases.

Challenges: Coaches were asked during site visit interviews about the challenges they had faced as a reading coach. Coaches were most likely to mention uncomfortable interactions with teachers and a lack of time to do everything expected of them. Most of the time, the resistance they encountered came from only a subset of teachers. Some of it was overcome as the year wore on, as teachers became more comfortable having coaches in their classroom, or more familiar with the core reading program and expectations of Reading First:

> *It was hard for teachers to let me in their rooms. I needed to learn to be careful and not to step on their toes. (Coach)*

> *It has been a year of transition, trying to get everyone on the same page and get past teachers' apprehensions. (Coach)*

> *It took time to win over the teachers so they could see the coach as a benefit, someone to make their job easier, rather than as a threat. (Coach)*

On the other hand, at several of the schools visited, the resistance the coach encountered was more serious and long-lasting:

Some teachers have been rather hostile. (Coach)

The hardest thing was resistance from the teachers to the program and to the principal trying to support what the district asked them to do. Some teachers were hostile, even unprofessional…the coach is leaving the school next year because of this. (Coach)

The other major challenge that coaches faced was insufficient time to accomplish all that was expected. One coach from a fairly large school noted that she was supposed to individually coach 21 teachers with different needs. Another said that because the school did not have an assessment coordinator, she was juggling both positions and lacked sufficient time to be in the classroom. Another said that the monthly trips to Phoenix for coach trainings meant she missed two days at the school and fell behind in record-keeping and other tasks; yet another echoed that thought, saying she was constantly "playing catch-up."

Perhaps one of the reasons that some coaches felt overwhelmed by the tasks confronting them, besides the novelty of the position, lay in problems with role definition. A few coaches complained that their jobs did not correspond to their original understanding of what the position would entail, and two were frustrated that they did not receive formal job descriptions until well into the school year.

Other difficulties that were mentioned included the amount of material there was to learn, and the challenge of providing feedback "diplomatically" to teachers (especially when these were former colleagues and friends). In addition, there were a slew of more individualized challenges, such as balancing work with family crises, disagreement with components of the core program, inadequate office space, and a sense of insufficient access to the level of technical assistance more available to schools in the Phoenix area.

Assessment Coordinators

To meet the data needs of the grant, the ADE designed a position called assessment coordinator. They envisioned the assessment coordinator responsibilities would include organizing data collection and management systems, overseeing and providing assessment, attending state trainings about assessments and bringing relevant information assessment coordinator for school staff. The vision for the role was not only to strengthen data collection and use of data at Reading First schools, but also to take on data responsibilities so that reading coaches had more time to work directly with teachers.

According to coaches, the majority (83%) of visited schools had hired an assessment coordinator. Over half of the assessment coordinator positions (58%) were part-time. Interestingly, part-time positions did not correspond with size of school; several of the largest schools reported they had part-time assessment coordinators while several smaller schools reported that the assessment coordinator position was full-time. In the four schools where there was no assessment coordinator, the reading coach took on most assessment responsibilities.

Assessment Coordinators described their main role as coordinating most or all aspects of DIBELS assessment. Among their many responsibilities, assessment coordinators listed the following tasks:

- Managing data systems
- Entering DIBELS results online
- Managing assessment teams
- Sharing administration of the DIBELS benchmark assessments three times a year
- Coordinating and participating in ongoing progress monitoring (e.g., assessing the lowest 20 percent of students every four weeks; assessing new students)
- Preparing data to share with staff (e.g., creating user-friendly charts or reports)
- Attending professional development provided by the state
- Providing professional development for staff about the assessments
- Communication with the coach, principal, and/or Reading Leadership Team

Some assessment coordinators reported that they also met individually with teachers to discuss their assessment results and/or attended regular grade-level meetings where data were discussed. Some assessment coordinators also reported that they were very involved in planning interventions and in some schools, assessment coordinators coordinated other assessments such as the DRA or SRI.

With multiple responsibilities and data from hundreds of students, it is not surprising that many assessment coordinators reported during site visit interviews that they felt they did not have enough time to accomplish all the things their job demanded:

> *The biggest challenge is feeling that I can't do what I need to do because of a lack of time. (Assessment coordinator)*

> *The biggest challenges are time constraints on getting and entering data. It is hard to keep up because there is a lot of student mobility at our school [lots of new students who need to be assessed]. (Assessment coordinator)*

> *We could be much further if there was more time. There is more that we want to do. There are so man different ways that we want to take the data apart. (Assessment coordinator)*

Some assessment coordinators felt that their work load would be more balanced next year because teachers would have more responsibility for administering the DIBELS and using DIBELS data.

A few assessment coordinators felt the biggest challenge had been working with staff that was resistant to the grant and/or the data:

> *I have taken on the burden of not being liked [because of the data] and have seen teachers' anger transferred to me. (Assessment coordinator)*

However, the majority of assessment coordinators felt their job had been rewarding and described many positive changes they had contributed to in Year 1:

> *The most rewarding aspect of my job has been seeing changes in the school and being part of it. It is exciting to see data make an impact. I have helped teachers be better [at using data] and have seen the kids succeed. (Assessment coordinator)*
>
> *The [new] visual representations of where the child is at in reading are very rewarding. Teachers are seeing the effectiveness of their work; seeing the progress made by students. (Assessment coordinator)*
>
> *I feel like what I do matters. I've seen students get excited about reaching benchmarks. And I've been able to pinpoint problems for teachers to address. (Assessment coordinator)*

Reading Leadership Teams

Schools were asked to establish Reading Leadership Teams (RLTs) that would bring teacher representatives together with the principal, coach, and assessment coordinator to oversee the implementation of Reading First. According to the ADE, RLTs should meet monthly and have established group norms and procedures for their meetings (for example, having an agenda and taking minutes, which should be shared after the meeting). The team should serve as a center for shared decision making about the school's and students' needs in reading. The RLT should create a schedule that ensures that the 90-minute reading block is in place and uninterrupted, and permits "double-dosing" (interventions) to occur during the school day. The team should regularly look at data together, analyze it and discuss ways to share it constructively with school staff.

Membership on the RLTs was fairly consistent across all visited schools. In all 23 schools, the principal and reading coach were identified as being on the RLT, as well as the assessment coordinator if the school had one. The majority of coaches reported having one teacher per grade level, although some schools had two or three per grade; and, in a few cases, there were no teacher representatives from some grades.

In most cases, RLT members did not include parents or paraprofessionals; however, in some cases a district staff member (23%) or Title I staff member (36%) was on the team. More than half of the schools (64%) had at least one special education teacher on the team.

The results from the implementation checklist suggested that, while RLTs became more established in Spring 2004, their leadership role was still evolving. For example, CRSs reported that RLTs fully monitored the school-wide progress of intervention plans at only eleven schools (17.5%) while twenty-four schools (38.1%) showed "scant" evidence of this role. There was also variation in response to items about RLT maintaining staff focus on reading goals and using group processes.

Table 8.7
RLT Items (Implementation Checklist)

	May 2004 Percent (n)			Change from February*
RLT revises/updates the ASIP to align with the school's Reading First plan.	88.7 (yes) (47)			+ 6 schools
RLT meetings occur at least once per month.	91.9 (yes) (57)			+ 5
	No/ Scant	**Some**	**Full**	
RLT uses group process guidelines/norms.	9.5 (6)	20.6 (13)	69.8 (44)	+ 17 schools
RLT prioritizes reading goals and maintains staff focus on the goals.	22.2 (14)	33.3 (21)	44.4 (28)	+ 15
RLT implements an appropriate school-wide intervention plan based on data.	30.2 (19)	36.5 (23)	33.3 (21)	+ 15
RLT monitors progress of the school-wide intervention plan and makes appropriate adjustments.	38.1 (24)	44.4 (28)	17.5 (11)	+ 15

*Number of schools that increased from "no" or "scant" evidence to "some" evidence or "fully implemented" from February to May. See text for discussion of limitations.

Similar to the implementation checklist, site visit interviews also revealed the evolving nature of RLT leadership functions; specifically that RLTs at only a few schools were performing most or all of the leadership functions envisioned by the ADE. Interviews revealed that in about half of the schools, the RLT served primarily as a vehicle to exchange information. Some examples include:

> *We're kind of the go-between connecting the leadership to our grade levels, we report back on progress monitoring, standards, etc. (Teacher on the RLT)*

> *It has been a great vehicle to get our voices heard and we have seen results. (Teachers on the RLT)*

The reported leadership functions reported in a few schools included engaging in shared decision-making, reviewing data together, and working together to ensure the fidelity of the program. For example, one school said their RLT made data-driven decisions; they reported using DIBELS results to drive instruction as well as to determine recipients of interventions. In another school, participants explained, "We have everyone on the team and we all make decisions together to reach Reading First goals." Another interviewee talked about how RLT meetings provided an important vehicle for teacher involvement in grant implementation, a place to "have their voices heard."

A few schools reported the role of the RLT had yet to be established. This resulted in few or no meetings before the site visit. One school reported, "We didn't learn until mid-year that we needed a team [so] we've only met three times." In another school, one participant explained why they hadn't met: "The teachers were not comfortable in allowing just one teacher from each grade level to speak on their behalf."

Communication, Collaboration and Support for Reading First

Levels of Support for Reading First

In reviewing a broad range of research studies examining the effectiveness of school reform efforts, Fouts (2003) noted that teacher attitude toward reform efforts was a common characteristic of the successful schools. For example, the Washington School Research Center (2002) summary of findings about the schools that had the most success in increasing student achievement found that:

> *A fundamental characteristic of these schools is that the majority of the educators are "on board" with the state reform efforts… the educators have all agreed, either because of philosophical belief, acceptance, or acquiescence, to move the school in a certain direction. A logical necessity of this agreement is the personal willingness of each teacher to give up long-held beliefs and practices at the school and classroom level. (p.22)*

Not surprisingly, teacher buy-in was noticeably absent in the schools whose students struggled most in the standards-based environment (McCarthy and Celio 2001).

Because of the importance of buy-in and a collaborative effort to create change, the surveys asked teachers and specialists about their level of support for Reading First. As shown in Table 8.8, 71.6 percent said the support statement was "very true" at pre-test; this declined to 63.7 percent at post-test. Levels of support were very similar across respondents regardless of their years of experience. However, data did reveal a larger decline in support among third-grade teachers compared to other grades; the percentage who believed this statement was "very true" declined from 68.7 percent to 53.1 percent over the year.

Table 8.8
Support for Reading First (Survey Responses)

I strongly support the instructional & other changes that [will] occur under the Reading First grant.	**Percent responding***	
	Pre-test	**Follow-up**
Very True	**71.6%**	**63.7%**
Somewhat True	**26.2%**	**31.8%**
Not True	**2.2%**	**4.5%**

*Includes 559 teachers and specialists who had both pre- and follow-up surveys. The actual "n" for pre-test and follow-up varies slightly due to missing cases.

Along the same lines, site visitors asked principals, coaches and assessment coordinators at Arizona Reading First schools to describe the levels of teacher buy-in and support for Reading First. Schools tended to fall into one of three categories. A few had either very high or very low support for Reading First, right from the beginning of the year. The largest grouping was in the middle, where teachers were divided in their support for the principles of the grant. Each of these groupings is described below.

Schools with High Initial Support for Reading First: In general, schools with high support for Reading First at the beginning of the year indicated that they maintained high levels throughout the year. These schools attributed their success to a variety of factors. For some, it was driven by teacher belief in the core reading program. Some schools with higher proportions of new teachers felt this was a benefit; as one principal explained, "Newer teachers are more welcoming to the coach and open to the program" than more experienced teachers. A few attributed their strong buy-in to "luck", acknowledging the "problems that other schools have had" regarding resistant staff.

Interviewees from one school explained that their school secured and maintained high support by laying a strong foundation and securing teacher involvement. From the beginning, administrators clearly delineated the process, expectations and options regarding the Reading First grant to all teachers. This included involving teachers in the grant application process, explaining the Reading First program to teachers in groups and one-on-one meetings both before and after the grant award, and providing an atmosphere that was receptive to teacher questions. Likewise, teachers were able to select the school's core reading program and were provided with a $1500 annual stipend for being a Reading First teacher, a gesture that "added an incentive and made them feel appreciated."

> *We probably approached this differently. Prior to finding out that we were selected to be a Reading First school, we made all teachers aware of what was on the application. Teachers were involved and started buying-in then. When we found out we were selected, we had a Reading First contract for the teachers. We met with each teacher individually and gave them the opportunity to ask questions and to decide if they wanted to move to another school if they couldn't make a commitment. Everything has fallen into place because we did our homework early on. (Assessment coordinator)*

Schools with Mixed Initial Support for Reading First: Most schools noted that support for Reading First was mixed – with many teachers "on board" and a few "hold-outs." Interviewees indicated that while some teachers "loved it" and were "excited," others were "still hesitant" or even "disgusted." A few schools cited that Reading First elicited strong sentiments on both ends of the spectrum; as one coach noted, "Boos and cheers from the side make up the loud minority."

For some, mixed support was driven by grade-level specific issues; for example, one school cited that resistance was concentrated among half-day kindergarten teachers who were concerned about the lack of time for other subjects when implementing the 90-minute reading block. Many others indicated that support was a function of the individual, with teachers displaying a wide variety of attitudes and quality of implementation.

Nonetheless, interviewees indicated that buy-in improved over the course of the year at almost every school. However, the majority of these gains were moderate; responses reflect a decline in complaints, rather than full-fledged support:

> *In the beginning, we would hear, "What about ELL," "Kids don't learn this way," and "Nurturing is what's best." Now we have seen more of the teachers asking how they can help, moving away from criticism to acceptance if not overt enthusiasm. (Principal)*

> *They are not fighting with us anymore. They are staying with the core. (Principal)*
>
> *It was a little slow at first, but there are not so many complaints now. (Coach)*

Interviewees cited several factors that helped increase support, the most frequent being increases in student performance as determined by teacher observation and/or DIBELS scores. In the words of one principal, "the DIBELS data are giving teachers positive strokes."

At many schools, strong principal leadership in reading was also cited as a positive factor, with the principal encouraging people to "get on-board," establishing the school's focus on reading, and awarding and celebrating Reading First successes throughout the year.

At other schools, the new core program and other reading materials helped foster buy-in. For example, one school reported that their school had not used a core program for many years, which had been difficult; teachers there eagerly embraced the new materials. At another, teachers found that the new materials enabled them to address the needs of lower-level students, who previously had been falling through the cracks.

Schools with Low Initial Support for Reading First: A few schools reported low levels of support for Reading First. While many schools struggled to obtain the support of each and every teacher, respondents from these schools indicated a more pervasive resistance among their staff. They reported that teachers did not like being told how to teach and were "combative," even displaying "overt hostility in meetings and some quite unprofessional behavior."

In general, those with poor initial support experienced continuing difficulties:

> *I must have heard teachers say, "Well, we didn't really want the grant in the first place," at least one thousand times. Resistance is a major issue. – Principal*

Interviewees noted several factors that hurt buy-in at their schools. The most frequently cited was the challenge of learning a new core curriculum. At a few schools, this was exacerbated by inadequate training or receiving materials late. For others, the content of the core itself was a sore spot; teachers found flaws and shortcomings with the materials or felt they were not appropriate to their students. One coach explained that when teachers' concerns with the new core program were not addressed, this exacerbated generalized frustration with the grant.

Several interviewees also touched on the idea that securing buy-in can be difficult during a start-up year. This occurs for a myriad of reasons, one of the most obvious being the natural resistance to changing established ways of doing things. Many noted that in the beginning of the year, they encountered a good deal of skepticism about the Reading First program and instructional model, or that teachers wanted to continue to determine for themselves how they would teach. In addition, the sheer learning curve for teachers – including a new core curriculum, new instructional strategies, new school structures, a range of meetings, and learning how to interpret assessment data – affected buy-in early in the school year.

Two survey items provide some possible insight into reasons for mid- or low-level support for Reading First. As shown in Table X, only 17.9 percent of teachers and specialist at survey pre-test said it was "very true" that they had a good sense of how Reading First would affect their teaching practices. Even fewer respondents (9.7%), said "very true" at follow-up, suggesting that, perhaps, some participants had been surprised by the expectations and were no longer sure what to expect in the future.

The second item in Table 8.9 indicates mixed beliefs about the importance of scientifically-based reading research on instructional practice; one-third of teachers found it "very true" at pre-test and follow-up; approximately 43 percent found it "somewhat true." The remaining 20 percent who did not find the statement true may have been among those who more strongly resisted Reading First in Year 1.

Table 8.9
Items Related to Support for Reading First (Survey Responses)

	Percent responding* Very true/Agree		Percent responding* Very true/Agree	
	Pre-test	**Follow-up**	**Pre-test**	**Follow-up**
I have a good sense of how Reading First will affect my/their teaching practice.	53.4%	44.3%	17.9%	9.7%
I strongly support the instructional & other changes that [will] occur under the RF grant.	**26.2%**	**31.8%**	**71.6%**	**63.7%**
I am convinced that scientifically-based reading research is a crucial tool to help teachers make decisions about their instructional practice.	42.7%	43.5%	35.6%	35.6%

*Includes 559 teachers and specialists who completed both pre- and follow-up surveys. The actual "n" for each item varies slightly due to missing cases.

Communication and Collaboration

Communication and collaboration among staff members is an important component of the Reading First program. Respondents indicated through survey items, the implementation checklist, and site visit interviews that many improvements were made in this area throughout the year and that they strongly supported the importance placed on communication and collaboration in their schools.

Specifically, survey responses revealed that over 93 percent of teachers and specialists expressed strong support for the value of collaboration among teachers by agreeing with the statement, "To be an effective teacher, it is essential that I have the opportunity to collaborate with fellow teachers." This support did not waver throughout the year.

Interviewees at every school also indicated that there was improvement in communication and collaboration as a result of the grant. The degree of improvement varied greatly across schools however; this was likely a product of how much "room to grow" they had as well as levels of staff support for Reading First and its program requirements, many of which directly encouraged new ways of staff speaking and working with each other. Respondents described several different structures and mechanisms that helped increase communication and collaboration; the three most commonly noted during interviews were meetings, assessment data, and professional development.

Reading-related Meetings

In interviews, by far the most frequently cited structure that increased communication and collaboration was the establishment of focused, purposeful meetings centered around reading. The strong majority of schools noted that they communicated about reading through the grade-level and Reading Leadership Team (RLT) meetings required by the Reading First grant.

Data from the implementation checklist on grade level meetings also indicated that most schools were using grade-level meetings as an avenue for communication and collaboration (see Table 8.10). Specifically the CRSs reported that in 86 percent of schools there was "some" or "full" implementation of regular grade-level meetings about reading instruction. Fewer schools, but still the majority, used grade-level meetings to discuss assessment data (75.2% "some" or "full").

Table 8.10
Communication and Collaboration (Implementation Checklist)

	May 2004 Percent (n)			Change from February*
	No/ Scant	Some	Full	
Teachers meet regularly in grade-level meetings and the time is used effectively to discuss, plan, adjust instruction and collaborate.	14.3 (9)	52.4 (33)	33.3 (21)	+ 2
K-3 teachers regularly discuss assessment data at grade level meetings to monitor progress toward benchmark goals.	23.8 (15)	47.6 (30)	28.6 (18)	+ 17

*Number of schools that increased from "no" or "scant" evidence to "some" evidence or "fully implemented" from February to May. See text for discussion of limitations.

Interviewees had a great deal of positive feedback about grade-level meetings. For many, grade-level meetings were the primary time during which collaboration took shape, where discussions happened and staff shared information and practices. In addition to formalizing the collaboration process, interviewees appreciated that they were able to draw in all staff rather than just those who tended to participate or dominate ("everyone is equal").

Another theme was the increased quality of discussion among teachers:

> *We are hearing professional discussions at grade-level meetings about what is working, what isn't working, and what we can do to help the kids. The principal and I sit with our mouths open because of the professional mindset that is happening at these meetings that hasn't been there before. (Coach)*

> *Reading First has had a big impact: constant discussion is now heard in the hallways. Reading has taken on a new sense of importance. (Coach)*

For most schools, this communication was new: as one coach noted, "We never met or talked about anything last year!" Among the few schools indicating that they had grade-level meetings in place before receiving the Reading First grant, virtually all noted that the focus, quality, usefulness and regularity of these meetings had improved over the past year.

Three of the schools visited reported that grade-level meetings at their schools had met with limited success, they had not quite "gotten off the ground" or failed to meet expectations for the level of "professional conversations" that coaches had anticipated. Future technical assistance from the ADE might target these schools to help entrench and maximize the effectiveness of their reading meetings.

Sharing of Assessment Data
Another structure frequently cited by interviewees as increasing communication and collaboration was the DIBELS assessment data. Looking at, sharing and using assessment data provided a launching pad for discussion, both among teachers and between teachers and parents:

> *Teachers have been looking over the data together; then they are able to communicate these scores to parents, who otherwise may wonder why their child cannot read. (Principal)*

> *It's a ripple effect, teachers want to go and share data. (Principal)*

> *I see them as a changed group. Instead of closing their doors, now they share. They talk about DIBELS results, what they mean and what they can do to improve them. Everything has been evolving. (Principal)*

Shared Professional Development
Another mechanism noted by interviewees as leading to increased communication and collaboration was shared professional development experiences. This included practices such as teachers observing each others classrooms or teachers even "trading classes to try different strategies." Others cited regular early release days as a time for teachers to come together for in-service training and sharing with colleagues. One effect of professional development that tends to assist communication and collaboration is the establishment of a common, shared language, something many teachers and coaches said they valued.

Teachers and specialists were surveyed about how helpful they thought classroom observations were. There was strong agreement that it was helpful to have a knowledgeable person observe and provide feedback. However, the agreement declined slightly from pre-test to follow-up.

Table 8.11
Helpfulness of Observations (Survey Responses)

It is extremely helpful to have a knowledgeable person observe my class and provide feedback.	Pre-test*	Follow-up
Agree or Strongly Agree	81.3%	77.1%
Neutral	**15.5%**	**17.5%**
Disagree or Strongly Disagree	3.1%	5.4%

*Includes 559 teachers and specialists who had both pre- and follow-up surveys. The actual "n" for pre- and follow-up ranges slightly due to missing cases.

Communication Between the Principal and Reading Coach

As laid out by state program staff, two-way communication between the principal and the coach is important for effective implementation. In most instances the coach and principal described each others' roles in similar terms. Some respondents specifically spoke of the collaboration between them, referring to a "common vision" and "healthy exchanges of ideas."

Site visitors observed very positive relationships between the coach and principal at most of the 23 schools visited. At over half of the schools, the coach and principal met daily or almost daily to collaborate on program implementation; at another seven, they reported meeting together at least weekly. While at two of the visited sites, the principal and coach had a tense or even adversarial relationship, most interviews attested to close collaboration and communication:

> *We talk ALL the time. (Coach)*

> *The principal and coach reported constant, daily communication about the implementation of Reading First. The principal reported that during and after classroom observations, she routinely asked the coach for clarification on what was observed and they shared perspectives on what needed to be strengthened. (Site visitor)*

> *The principal shared perceptions of observations with the coach, and the coach shared concerns with the principal, who backed her up 100 percent. The principal and staff recognized the coach as a master teacher; the principal valued her expertise and therefore conferred regularly with the coach on implementation issues. (Site visitor)*

CHAPTER 9
DATA AND ASSESSMENT SYSTEMS

Highlights

- Schools reported that they used the DIBELS assessment, as well as core program assessments to "inform instruction," by which they primarily meant screening, progress monitoring, and assigning students to groups.

- Advantages of the assessment systems were reported to be that data served as a basis for teacher collaboration and planning and communication with parents about their child's progress.

- Most schools appreciated and valued the DIBELS assessment system, particularly for the timeliness of the information they got back. At a few schools, there were concerns about the applicability of DIBELS to new ELL students, to older students who read at very low levels, or about contradictory results from different assessments.

- Because the newly established assessment systems were generally valued, schools that looked ahead were already concerned about sustaining them beyond the life of the grant. Long-term sustainability of an assessment system might be an area for the ADE to begin exploring with districts and schools well in advance of the end of the Reading First project.

Data and Assessment Systems

Reading First schools were expected to review, discuss, and use data from the DIBELS and other student assessments in various ways. According to the ADE, data review were to make up a part of cross-grade and grade-level collaboration and be a focus of the Reading Leadership Team. The reading coach and assessment coordinator should help make grouping recommendations for intensive interventions based on data.

According to the ADE, data should also be used in progress monitoring for all students, although particularly for students in need of interventions. Progress monitoring should include data from core program unit and theme assessments, as well as progress monitoring assessments from the DIBELS.

In order to facilitate the establishment of a systematic approach to assessment, detailed DIBELS trainings were held in fall 2003. Assessment teams were required to attend and participate (see professional development section for details). The DIBELS assessment was administered to all K-3 students three times during the 2003-04 school year, and data were submitted to the University of Oregon database for project level analyses.

Implementation and Use of Assessment Data in the Reading First Schools

All Arizona Reading First schools adopted the DIBELS as a common outcome measure for the project. During site visits, coaches listed 36 different assessments that they used in addition to the DIBELS, including assessments associated with core programs (19.4% of those cited) and phonics and fluency assessments associated with supplementary programs such as Read Naturally (27.8%). The remaining assessments (41.7%) were a variety of assessments cited by only one school each. About one in ten visited schools reported using no other assessments.

Systematic implementation

Once assessments were identified, the expectation was that Reading First schools would systematically implement systems to collect and analyze assessment data for the purpose of improving instruction, particularly for struggling readers. The implementation checklist, administered in February and May, provided information about the CSR's perceptions concerning the extent to which assessments were systematically implemented in Arizona Reading First schools. By May 2004, the majority of schools (82.5%) were rated by CRSs as having fully implemented the DIBELS and established a system for collecting and analyzing those data in grades K-3. The frequency of identification and use of other assessments (diagnostic and core program assessments) also increased from February to May, but lagged slightly behind DIBELS implementation (Table 9.1).

Table 9.1
Assessment Systems (Implementation Checklist)

	May 2004 Percent (n)			Change from February*
	No/ Scant	Some	Full	
A system is in place to administer, score, report, share and analyze DIBELS for grades K-3.	1.6 (1)	15.9 (10)	82.5 (52)	+3 schools
A system is in place for identifying, assessing and monitoring at-risk students, or those in need of intensive intervention.	3.2 (2)	20.6 (13)	76.2 (48)	+5
Diagnostic assessments have been identified and selected for use with identified students and are administered accurately and in a timely fashion.	12.7 (8)	31.7 (20)	55.6 (35)	+15
Assessments within the core reading program are identified and selected for use for each grade level K-3 and are administered accurately and in a timely fashion.	17.5 (11)	34.9 (22)	47.6 (30)	+11

*Number of schools that increased from "no" or "scant" evidence to "some" evidence or "fully implemented" from February to May.

Survey responses from teachers, specialists, and coaches, collected in the fall and spring, closely mirrored the implementation checklist results related to the level of implementation of assessment systems in Reading First schools. By spring, 94.3 percent of respondents agreed or strongly agreed that a system for collecting and storing reading assessment results was fully implemented (an increase from 78.2 percent in February).

In addition, the survey results also documented increased uniformity in the use of assessment instruments from fall to spring. When teachers, specialists, and coaches were asked whether teachers used different reading assessment from other teachers, depending upon experience or preference, 25.1 percent *strongly agreed* that this was the case in the fall. By spring, this level of agreement dropped to only 10.7 percent.

Uses of assessment data

In order to gain additional insight into how the assessment data were actually utilized in the schools, site visit interviewers asked respondents to briefly describe how data were used to look at patterns across the school and what they perceived to be the strengths related to their schools' assessment system. The most common response to both of these questions was that "data drive or inform instruction." These words were used so frequently that they seemed to be almost a slogan. Asked to explain more, most frequently, staff interpreted "driving or informing instruction" to mean that data were used to: (1) group students for instruction; (2) determine which students needed interventions; and (3) monitor student progress in interventions.

> *The major use of data was to identify students with the largest deficits…This meant that teachers worked with students at emerging level and tutors worked with very low deficit kids during small group intervention time. (Assessment Coordinator)*
>
> *We look at it for progress monitoring and to identify reading difficulties…it drives instruction. (Principal)*
>
> *We use the results to place kids in classrooms and it helps (identify) students for interventions (Principal)*

In addition to site visit interviews, survey data were also collected to determine participant levels of agreement with statements describing teacher/specialist collaboration, communication, and interpretation of assessment results. The survey data, presented in Table 9.2, were consistent with interview data and indicated that there was a substantial increase from fall to spring in K-3 teacher-collaboration concerning assessment data within grade levels. For example, those agreeing or strongly agreeing that "teachers sit down at least quarterly to review assessment data," increased from 55.4 percent of respondents in the fall to 89.2 percent in the spring. Across grade levels, there was a similar increase—from 46.6 percent to 76.1 percent.

Table 9.2
Assessment Items (Survey*)

Item	Somewhat True/ Agree		Very True/Strongly Agree	
	Pre-test	**Follow-up**	**Pre-test**	**Follow-up**
The K-3 teachers sit down together at least once a quarter to review reading assessment data in grade level groupings.	26.0%	23.3%	29.4%	65.9%
The K-3 teachers sit down together at least once a quarter to review reading assessment data across grade levels.	25.7%	32.3%	20.9%	43.8%
Teachers/specialists use one or more of the same assessments several times a year to monitor student progress	21.4%	7.8%	71.7%	91.2%
Teachers regularly share and discuss assessment results with other teachers in order to evaluate/plan.	41.9%	28.8%	44.7%	66.9%

*Includes 618 teachers, specialists, and coaches who had both pre- and follow-up surveys. The actual "n" for pre-and follow-p varies slightly due to missing cases; percentages are based on number responding

Similarly, nearly 96 percent of respondents strongly agreed that teachers regularly shared and discussed assessment results with colleagues for planning and evaluation. And, by spring, participants were in strong agreement (91.2%) with the statement that teachers/specialists use assessments several times per year to monitor student progress.

Strengths of the assessment systems
Interview data uncovered several advantages perceived by Reading First participants. One advantage cited was the DIBELS database maintained by the University of Oregon that provides instantaneous charts and graphs of the data entered online. Several interviewees mentioned that it was very beneficial to be able to access assessment results quickly:

> *The major benefit has been getting the results back early...so I can report them to teachers (Assessment Coordinator)*

> *We valued the instant, reliable feedback written in a narrative style that is easy for staff to read with a variety of graphs. (Assessment Coordinator)*

Additional advantages of having and using data in the schools emerged in many of the interviews. One prominent advantage often referenced was the use of data as a basis for teacher discussions, planning and professional development. Another advantage cited by respondents was that results served as a way to communicate with parents about children's progress. The examples below illustrate the thoughts of many:

> *Teachers talk about the data and use it for planning (Reading Leadership Team)*

> *Data are used for cross-grade impact—teachers talking more. (Assessment Coordinator)*

> *I look at [data from] individual teachers to see where we need more training. (Principal)*

> *Data tell us who needs help and we share it with the parents. (Principal)*

Survey data were consistent with the interview data concerning the use of assessment data to communicate with parents. On the spring survey, nearly 90 percent of respondents *agreed* or *strongly agreed* with the statement: "Assessment results, besides the state benchmark tests, are shared with/explained to parents."

A few interviewed staff expressed appreciation for the strong leadership and support they received from principals and assessment coordinators. One staff member appeared to be especially grateful for the support of the assessment coordinator:

> *The main strength of the assessment system is the assessment coordinator herself—conducting meetings, summarizing and sharing information. And the principal has been a key in her leadership during this process.*

Finally, a strength of the assessment system identified by a few respondents was the effect of sharing results with students on student motivation. An assessment coordinator reported that "…students are motivated and want to see how they did before and after assessment." A member of a school Reading Leadership Team echoed this sentiment and pointed out that "the kids' drive has increased due to the (assessment) goals set forth in the program."

Assessment-Related Challenges

Interviewed staff members described several challenges and concerns related to their assessment systems and how they used data. First, there was wide-spread concern that there had not been enough school-level training in the DIBELS and other assessments. Some interviewed staff expressed a strong interest in more information about how to use and interpret the data, particularly as a way to increase teacher involvement in the assessment process:

> *Teachers need training in the DIBELS so they understand more about what it is; and we also need more involvement. (Principal and Teacher)*

> *Staff need help in understanding assessment broadly and see how it fits together (Reading Coach)*

> *DIBELS training and data analysis should be an ongoing professional development focus at the school. (Reading Coach)*

There was an equally widespread concern about the time required for assessment, including analysis and application of data. Some participants voiced concern that there were too many tests, some of which were perceived to be redundant.

> *I have to administer five other assessments besides the DIBELS. This is just too much! (Teacher)*

> *There are a lot of assessments going on. Some are repetitive, and the DRA is not reliable. (Principal)*

> *A major challenge we face is not having enough time to keep up with the DIBELS testing and progress monitoring. We need help! This is a huge job. (Assessment Coordinator)*

A few schools expressed concern with the use of assessments with English Language Learners (ELL). They questioned whether or not they could use a student's native language for giving instructions prior to an English-only assessment:

> *As far as assessment, the assessor needs to be careful while listening to ELL students because they pronounce words differently and we don't give credit for words they don't pronounce correctly – you have to make a professional decision if that counts…Some items I allow them to retell in Spanish even if I can't count it, at least I can get more information on them. (Assessment Coordinator)*

> *We need something other than the DIBELS for …ELL students, especially coming from Mexico at the end of the year." (Coach)*

Some respondents were more specific in their criticisms of the DIBELS. An assessment coordinator at another school wondered about the applicability of the DIBELS to all students. For older students who read at very low levels, he felt that the oral reading fluency assessments were not sufficient and wanted to have norms on the phonemic awareness and phonics measures usually given only to students in kindergarten and first grade.

Other issues that were raised included concern about potentially inaccurate or conflicting results, and lack of emphasis on comprehension. Some of the participants' ideas are exemplified in the following examples.

> *For us, the DIBELS data does not match the core program assessment information—particularly for fluency. (Reading Leadership Team)*

> *The teachers had concerns about the lack of focus on comprehension. (Site Visitor)*

Finally, a few of respondents were concerned about sustainability. One principal voiced concern by pointing out that "….there is also the issue of what will happen once the funds are gone and there remains nobody to assess their (teachers') kids." A member of the Reading Leadership Team and an Assessment Coordinator both wondered, "…what will happen after three years…how will the program be sustained?"

CHAPTER 10
Instruction

Highlights

Core Reading Programs

- Overall satisfaction with the different core reading programs at the visited schools was high, regardless of which program was adopted. Teachers and specialists rated their core programs high in areas related to the five essential components such as providing explicit instruction in comprehension strategies and exposing students to vocabulary in a variety of contexts.

- Although overall satisfaction was high, most interviewees pointed out gaps and flaws in their core programs. For example, they reported that some core program manuals were not user-friendly for teachers; writing was not integrated in the program; materials were too difficult for ELL students; or certain components such as comprehension were stronger in some grades than in others. Also, some schools complained about difficulties in obtaining all the materials the publishers had promised them.

- Schools knew that "fidelity to the core program" was important to state project staff. Many schools tended to interpret this very strictly, as a charge to teach only and exactly what was in the book. Some schools articulated a comfortable balance between the scope and sequence of their core program and thoughtful adaptation to student needs, while a few schools appeared to adopt a looser interpretation of fidelity ("the spirit, not the letter, of the program"). The core program was used in almost all observations conducted by site visitors.

Instruction

- Site visitors observed instruction in all of the five essential components of reading. Kindergarten lessons tended to focus on phonemic awareness and phonics. In first grade, these were also the primary areas of instruction, although work in fluency and comprehension were observed as well. In second and especially in third grade, comprehension was the most commonly observed focus of instruction, followed by vocabulary and fluency.

- Phonemic awareness and phonics lessons tended, for the most part, to be fast-paced and to promote student engagement. In the area of vocabulary instruction, site visitors witnessed both highly engaging, active lessons, and slower, less interesting reviews of vocabulary words. In the area of comprehension, it was more common to see teachers asking students to answer simple recall questions than to think more deeply about the meaning of text. These observations mesh with statements from coaches, who often reported that vocabulary and comprehension instruction were the areas they felt needed most improvement in their schools.

- Flexible grouping is an area that was rated low in the implementation checklist. Additionally, in at least ten percent of observations, site visitors noted that lessons were either clearly too

difficult for students or too easy, with teachers continuing to belabor concepts that students had mastered and were bored with practicing.

- Because observations were comparatively short, one-time visits, it cannot be assumed that failure to observe a particular strategy meant that teachers never used such a strategy. What observations do suggest, when combined into a big picture, is the degree to which certain practices have become customary in some classrooms. For example,

    - In about half of observed lessons, teachers regularly monitored student understanding and adjusted the pace of instruction to fit student needs.
    - In about the same proportion of observations, site visitors observed that the teacher provided clear and appropriate feedback to students as they worked.
    - Explicit modeling was noted in about half of all observations.
    - In terms of getting the most use possible out of available instructional time, site observers noted some examples of excellent classroom management that moved students efficiently from one task to another, but in many cases slow transitions or the failure to have materials ready ahead of time meant there was wasted time in the classroom.
    - There were also many instances in which large numbers of students sat listening (or not) while only one student read; many teachers did not make use of think-pair-share, partner reading, or other strategies that would have meant more students were practicing at the same time.

Interventions

- During the first year of implementation, state project staff placed primary emphasis on implementation of the core program, leaving work on the establishment of an intervention system for the second year. Not surprisingly, items related to intervention strategies were among the lowest-rated items on the implementation checklist. In addition, thirteen of the 23 visited schools had not purchased any supplementary or intervention materials.

- Some schools had built an intervention program. Often they reported using materials attached to their core reading program, with varying degrees of success. A few schools had well-developed systems, complete with appropriate supplementary materials and regular progress-monitoring.

- Many schools struggled with the scheduling and staffing demands of putting together enough interventions for all the students who needed it. Schools that were able to draw on volunteers or paid outside tutors or had teachers work extra hours to provide interventions were very grateful for the additional staffing.

- School staff requested training and technical assistance to increase buy-in for creating a schoolwide intervention plan, to learn to better interpret and use DIBELS data, to better target interventions to specific needs, and to determine which materials best serve which types of purposes.

Meeting the Needs of ELL Students

- Almost half of the schools visited reported that English language learners (ELLs) made up a majority of their student population, and most of the other schools had at least 20 percent of their student population learning English. Consequently, addressing the needs of ELL students was an important issue to schools.

- At the end of Year 1, about two-thirds of teachers, specialists, and coaches said it was "very true" that they could describe and model effective strategies to work with ELLs; most of the remaining respondents responded that this was only "somewhat true." This corresponded with findings from survey interviews; two-thirds of the observed teachers were able to describe very specific modifications they made in their teaching in order to support ELL students. Most commonly, these included bringing in visuals to support concepts taught in class and incorporating physical experiences into the lesson.

- Opinions about how well the core reading programs served the needs of ELL students varied tremendously. Many teachers relied on supplemental ELL materials that came with the core program and found these to be useful. Still, teachers, principals, coaches, and districts all expressed a need for more information and more support on working effectively with ELL students.

Characteristics of Instruction

The previous sections of this report documented the work of Arizona Reading First in the professional development for teachers, principals, coaches and assessment coordinators as well as the creations of collaborative structures and systems to use data. The ultimate purpose of that work is to influence the content and quality of reading instruction at the classroom level. According to the original Reading First proposal, material presented at monthly coaches' and principals' trainings, and an extended interview with the state project director, the Arizona Reading First classroom, fully implemented, includes the following characteristics:

- Students receive at least 90 minutes a day of uninterrupted reading instruction that relies on the core reading program adopted by the school.

- Instruction follows the scope and sequence of the core reading program, which has been developed, based on scientific research on reading and addresses the five essential components of reading: phonemic awareness, phonics, fluency, vocabulary development, and comprehension strategies.

- Instruction in the five essential components should be both systematic and explicit.

- Both whole and small group instruction should be utilized; small group instruction provides an opportunity to target instruction to specific student needs, identified through appropriate assessments.

- Students should receive both grade-level and instructional-level instruction.

- Efficient classroom management is needed, and routines should be established and functioning in order to maximize the amount of time available for instruction.

- Teachers should use a variety of approaches to ensure active student participation and high levels of engagement.

Evaluators collected information about instruction through survey items, the implementation checklist and during interviews and classroom observations in a subset of Arizona Reading First schools.

The majority of teacher, specialist and principal survey respondents (94.5%) believed that instruction had improved over the year. However, this growth is balanced by evidence from the implementation checklist and classroom observations which suggests that there is still room for substantial growth in the area of instruction. This section describes these findings.

Use of the Core Reading Program

All Arizona Reading First schools were required to adopt a core reading program for grades K-3. The majority of teachers, specialists, and coaches who responded to survey questions strongly agreed ("very true") that their core program were strong in areas related to the five essential components. For example, as shown in Table X, over 80 percent responded that it was "very true" that their Reading First program integrated word recognition, fluency, and vocabulary skills; integrated the essential components of reading; and exposed students to vocabulary in a variety of contexts (in most cases, the remaining respondents believed these statements were "somewhat true"). Fewer respondents, but still the majority, believed that the reading program included small guided reading groups or provided reading practice with texts matched to individual reading levels.

Table 10.1also indicates that respondents were much less likely to believe that the reading program they used prior to Reading First had these qualities; "very true" responses ranged from 44 to 63 percent for previous reading programs compared to a much higher range of 65 to 88 percent for Reading First core programs

.

Table 10.1
Previous/Current Reading Program (Survey Responses)

	Percent Responding "Very True"*	
	Previous Reading Program (pre-test)	**Reading First Core Reading Program (follow-up test)**
The reading program integrates word recognition, fluency, and vocabulary skill.	63.2%	87.5%
The reading program integrates the essential components of reading.	43.6%	86.3%
In the reading program, teachers expose students to vocabulary in a variety of contexts.	58.7%	82.2%
In the reading program, phonemes are linked with letters as soon as students understand that the letters represent sound.	52.4%	77.3%
In the current reading program, students are provided explicit instruction in comprehension strategies, such as the use of graphic organizers, making connections and inferences, & citing evidence for their ideas.	49.0%	72.5%
In this reading program, students practice fluent reading using texts that have been matched to their individual reading level.	57.8%	74.1%
Part of the reading program includes students working in small guided reading groups, repeatedly reading the same text.	49.0%	64.9%

*Includes 618 teachers, specialists and coaches who had both pre- and follow-up surveys. The actual "n" for each item varies slightly due to missing cases.

About half of the 23 schools that were visited by evaluators had chosen Voyager as their core reading program; five schools had chosen Houghton Mifflin and four had chosen McGraw Hill. Success for All, Harcourt Trophies, and Harcourt Brace were chosen by one school each.

During site visits, all visited schools except one reported that they liked their core reading program and almost all teachers (95.8%) observed by site visitors used the core program for all or the majority of the lesson. Several reading coaches described how the core program provided a good overall structure and sequence for their reading program that had not existed before Reading First:

> *The core is sequential, specific, and spirals through the grades so it fits together well. Before [we adopted this] core program…there were huge gaps in our instructional program. The core has given teachers good structure. (Coach)*

Many coaches and members of Reading Leadership Teams also described specific things they appreciated about the core program such as decodable books, intervention materials, specific components such as comprehension or phonics, and teacher-friendly manuals.

Although overall satisfaction was high, most interviewees pointed out gaps and flaws in their core programs. For example, they reported that some core materials were not user-friendly for teachers; writing was not integrated in the program; materials were too difficult for ELL

students; or certain components such as comprehension were stronger in some grades than in others. For example:

> *The tasks in the core program are too difficult and sometimes vocabulary and stations are only loosely connected with the content of the core. It is not good for our ELL students because it does not focus enough on vocabulary. (Coach)*

Because the core programs were new to staff in most schools, teachers and paraprofessionals invested a large amount of time and effort to learn the programs. Coaches and teachers reported that the learning curve was steep and the amount of material to learn and prepare was a challenge. Several schools also reported that they had trouble getting all of their core materials, which created a rocky start to the year. A few schools were unhappy with the training provided by the core program publisher.

Fidelity to the Core Program

One of the messages communicated by the ADE to Reading First schools was the importance of using the chosen core reading program with fidelity. Once in the field visiting schools, evaluators noticed a range in the way that school interpreted that message, so to clarify the vision of fidelity, evaluators asked the statewide director of Arizona Reading First to explain the ADE's understanding of the term:

> *Fidelity is following the scope and sequence of skills… following the teachers' guide to ensure systematic and explicit instruction.*
>
> *It does <u>not</u> mean that every teacher needs to be on the same page on the same day. That's not responsive to student needs. But on the other hand, grade level teams should plan together and keep a pace appropriate to kids. This means the grade as a whole should be at about the same place - for example on the same unit if not the same story. There is a little bit of breathing room to adjust to classroom learning.*
>
> *Teachers are not expected to "read the script no matter what." This is where professional judgment comes in, where teachers make informed decisions about pacing. There should be a chance for teachers to collaborate with other teachers to figure out what works best.*

There was an acknowledged tension between following the core program, with its carefully defined sequential and systematic introduction of sounds, letters, and skills on the one hand, and adjusting instruction to student needs on the other hand. In this first year of implementation, the state chose to emphasize working closely with the core program and utilizing the teachers' manuals in order to communicate the centrality of the core program in reading instruction. According to the statewide director:

> *We didn't want teachers to think that the core program was not central to instruction, to think you could just skip what you wanted to or substitute in your own program. Then*

> *everyone would deviate. We tried to say: stick to the core in terms of the content you deliver; adjustments should be in pacing and grouping.*

The schools that were visited during Year 1 varied in their understanding of this message. During site visits, evaluators asked coaches directly to explain how they understood fidelity. Most understood the message of using the core program as it was designed, sometimes interpreted quite strictly:

> *Fidelity means you have to stick to it. All teachers have to use the core the way it was written.*

> *Fidelity means delivering the core program as it was written. Don't skip any elements. Cover the "Big Five" every day. Do what is in the book. Our district allows some enrichment in student engagement strategies but not in content. There is no room this year for teacher judgment about what to omit and what to include. We need consistency from school to school now.*

> *It means you stick absolutely to the core program and do not bring in any other curriculum.*

> *Fidelity is teaching all students for 90 minutes at their grade level, straight out of the book and teachers' manual. Everyone should be on the exact same page on the same day. This is a mandate from our district. Within the 90 minutes, teachers are not supposed to speed up or slow down according to student needs. Those adaptations can only occur outside the 90-minute block.*

A few schools seemed able to articulate the balance well:

> *Fidelity is following the core program as closely as possible while also meeting student needs.*

> *At the start, fidelity was using the core program during the block, regardless of the kids' levels. And that is what teachers did. It was as if teachers forgot that they had effective instructional strategies and the freedom to apply them. Now fidelity for us includes teaching skills and strategies to students, using the core, and delivering instruction in a manner that is appropriate to all students.*

In a few cases, schools appeared to allow themselves perhaps somewhat greater leeway than the ADE had in mind. For example, one coach reported:

> *It is the tenth year we are using the program, so we follow the spirit, not the letter, of the program.*

Asked the degree to which they thought teachers were actually implementing the core reading program with fidelity, coaches reported widely varying estimates, ranging from 65 to 100 percent of teachers and averaging 88 percent.

Teaching to Students' Instructional Level

The ADE envisioned Reading First schools providing students with both exposure to grade level material and targeted instruction at their appropriate level. This combination could be, and in practice was, delivered in different ways. For example, in Yuma and Somerton districts, students spent 90 minutes in their regular classroom receiving instruction at their grade level. Later in the day, they received another 60 minutes of instruction at their own reading level. More commonly, schools that were visited reported that they grouped students by their instructional needs within their regular classroom and used the core reading program at grade level with adjustments according to student level.

Prior to receiving Reading First grants, some schools already grouped students by level for reading instruction, but several schools reported that they had not grouped students at all prior to Reading First.

The classrooms observed by evaluators were predominantly self-contained classes (84.9%), that is, all students remained with their regular classroom teacher. Another way to deliver differentiated instruction is through "walk-to-read," in this system students leave their regular classroom and walk to a reading lesson with other students at their instructional level. Two of the 23 visited schools reported using "walk-to-read" to deliver targeted instruction. The average class size (students present on the day of the observation) was 19 students; only three observed classes had more than 25 students.

It was clear from implementation checklist results that the use of flexible grouping (small, changing groups of students made up of students with similar needs) is an area for further training. As shown in Table 10.2, items related to flexible grouping were rated very low compared with other items on the checklist. At the end of Year 1, 10 or more schools showed "no" or "scant" evidence of using and managing flexible grouping effectively; the majority of schools showed "some" but not "full" evidence in these categories.

Table 10.2
Flexible Grouping (Implementation Checklist)

	May 2004 Percent (n)		
	No/ Scant	**Some**	**Full**
Teachers use assessment data to determine flexible groups for additional instruction.	16.1% (10)	56.5% (35)	27.4% (17)
Teachers manage flexible grouping effectively so that all students benefit.	20.6% (13)	61.9% (39)	17.5% (11)
Teachers use flexible grouping to deliver additional instruction to students as needed.	23.8% (15)	50.8% (32)	25.4% (16)

In classroom observations at the 23 visited schools, small group work where teachers worked (six or fewer students) was observed about 20 percent of the time. During approximately two-thirds of the total time site visitors spent in classrooms, teachers were teaching to the whole

class. Because site visitors saw these classes only on a single day, however, it is not possible to know whether this accurately represents the ratio of whole group to small group instruction that actually took place.

Without knowing the learning background and achievement results of each child, it was difficult for site visitors to assess whether or not teachers were teaching at each students' instructional level. Nevertheless, it was clear to observers in approximately one out of every ten observations, that the instruction provided was not at the students' level. In some instances, the lesson was clearly too difficult for students, as in the example one site visitor recorded:

> *The teacher introduced the book Howdy Clown and asked one student to start reading. The student had a hard time reading the first page. The teacher told her to sound out the words and helped her.*
>
> *Students took turns reading the story, but many had a hard time. They were only reading very slowly, with help from the teacher. Students in the group who were not reading aloud were looking at other pages or not following along.*

In other observed classes, site visitors indicated that the material in the lesson had clearly already been mastered by students because they had no problems with any of the responses, and were sometimes bored by the ease of the lesson.

Monitoring Understanding

Site visitors were also asked to indicate the degree to which teachers monitored student understanding and used the information to adjust their lesson. Monitoring understanding might include calling on students randomly, asking questions, checking through repetition, or other strategies. In about half of all observations, site visitors indicated that teachers regularly checked for understanding in a variety of ways and used the information to adjust the content and pacing of their lesson. For example, in this kindergarten class with a focus on changing one sound in simple consonant-vowel-consonant words, the teacher checked that students understood the words used as examples and slowed the lesson down when needed:

> *Teacher: Bug. Let's break it down - /b/ u/ /g/. Find the letters.*
>
> *Students use letter boards to find the letters in the word.*
>
> *Teacher: Let's sound it out.*
>
> *Teacher and students: /b/ /u/ /g/.*
>
> *Teacher: Ok, give me five. [sentences with at least five words using the word "bug".]*
>
> *Student: I saw a bug crawling on the ground.*

Teacher: I saw a bug crawling on the ground. Ok. Let's change bug to rug…

[lesson continues in this pattern, using "run", then "tug"]

Teacher: Give me five.

Student: I have a tug.

Teacher. Maybe we need to figure out what tug means. Does anyone want to tell us what it means?

Another student: Carry something heavy.

Teacher: No that would be "lug". What does it mean to tug?

Another student: To pull.

Teacher: That's right. It is another word for pull.

In about 40 percent of the observations, the site visitors observed teachers checking for understanding, but moving on in the lesson without incorporating what they learned from these checks. A few observers indicated that the teacher rarely or never checked for student understanding or did so in a more superficial way (e.g., "Understand? Okay…"). These findings correspond with the majority of County Reading Specialists (74.6%) who indicated on the implementation checklist they had found "some" but not "full" evidence that teachers checked for understanding to make instructional decisions.

Feedback to Students

Site visitors indicated that approximately half of observed teachers provided clear and direct feedback to students during the observed lesson. In one kindergarten class, the site visitor recorded the following example:

The teacher directed students to take out their alphabet rainbows and start putting certain letters on the rainbow mat.

Teacher: "Oh, Sadie has a good start. She remembered about "b" and "d". Are you putting capital or lower case, Sadie?"

Sadie didn't reply. The teacher waited.

Teacher: Can someone else help Sadie? Is she putting capital or lower case letters?

Student: Lower case.

These exchanges continued, moving from student to student. The teacher usually praised their correct answers.

(Later) Teacher: Can you spell 'jet'?

Student: J-N-T.

T provided feedback to help him correct the answer. Teacher: "Put your finger on the first sound, the second sound, the third sound. Now slide the word, jet. Good!

In the above example, the teacher not only gave feedback that provided positive reinforcement, but also gave appropriate feedback when a student had an incorrect answer. In about 40 percent of observations, the site visitors observed teachers providing some feedback, though it was not always clear or frequently provided. Furthermore, according to site visitors, one in ten observed teachers provided little feedback to students or provided corrections that were not really helpful. For example, a third-grade teacher listening to students read "round robin" style quickly supplied any word students struggled over without either encouraging them to figure it out nor explaining what it was they did not understand. Ten minutes into the lesson, the site visitor noted that students had stopped trying to read difficult words and simply waited for the teacher to provide them.

Coverage of the Five Essential Components

Most of the data evaluators analyzed related to coverage of the five essential components came from classroom observations. Each of the 74 observations was divided into four five-minute blocks for a total of 257 five-minute blocks (some observations were only three blocks due to lack of transition time). During each five-minute block, the observer recorded one or two focus areas of the teachers' instruction.

Evaluators were most likely to observe phonics and comprehension instruction during the time they were in classrooms (See Table 10.3). Phonemic awareness, fluency, and vocabulary instruction were less commonly observed. In a small but noticeable percentage of observed blocks, the instruction was in another subject, or directions and transition were the major focus during the five minutes.

Table 10.3
Focus of Instruction in Observed Classes

	Percent of blocks*
Phonemic Awareness	19.8%
Phonics and Decoding	30.7%
Fluency	21.8%
Vocabulary	20.2%
Comprehension	40.5%
Other Subject	13.6%
Directions and/or Transition	7.4%
No Instruction	0.8%

*Percent adds up to more than 100 because blocks could have more than one focus.

The different areas of focus were not evenly distributed across grades. As Table 10.4 illustrates, lessons in second- and third-grade classrooms were most likely to focus on comprehension, with considerable work on vocabulary as well, while kindergarten and first-grade lessons often focused on phonics or phonemic awareness. These grade-level differences are discussed in the next sections.

Table 10.4
Percentage of Observation Blocks Devoted to
Each of the Five Essential Components, By Grade

	Grade Level			
	K	1	2	3
Phonemic Awareness	37.7%	32.4%	9.2%	1.5%
Phonics and Decoding	62.3%	44.6%	23.1%	0.0%
Fluency	15.1%	23.0%	18.5%	29.2%
Vocabulary	18.9%	12.2%	21.5%	29.2%
Comprehension	11.3%	27.0%	52.3%	67.7%

Note: Columns add up to more than 100 because more than one area could be chosen as a focus.

Phonemic Awareness

Phonemic awareness or the ability to recognize and manipulate sounds within words, the focus of instruction in about 20 percent of observation blocks. In order for students to learn that written letters signify sounds that can be combined to form words, students must first be able to recognize individual sounds (phonemes) within words and realize that combining sounds in different ways creates different words. Students learn to blend phonemes into words (so that /c/ /a/ /t/ blended together becomes "cat") and to segment words into their component phonemes (the sounds in "bat" are /b/ /a/ and /t/).

Explicit instruction in phonemic awareness is most commonly provided to younger students, in kindergarten and early first grade, although some work on phonemic awareness may continue on into higher grades as well if diagnosed as a specific need. In practice, there was still some work on phonemic awareness and phonics observed at the second-grade level, but almost none at all by the third grade (see Table 10.5).

According to survey respondents, there was a substantial increase in the percentage of teachers, specialists, and coaches who said it was "very true" that teachers focused on only one or two phonemic awareness skills at a time (from 34.6 percent at pre-test to 60.2 percent at follow-up). "Very true" responses to a statement about using explicit and systematic instruction in phonemic awareness also increased from 34.5 percent to 72.5 percent over Year 1 and "very true" responses to "Usually, teachers begin with auditory activities" increased from 47.10 percent to 75.3 percent. However, about one in four respondents did not believe these statements were "very true" even at follow-up.

Table 10.5
Instructional Practices Related to Phonemic Awareness
(Survey Responses)

	Percent Responding "Very True"*	
	Pre-test	**Follow-up test**
Teachers tend to focus on only one or two phonemics awareness skills at a time.	34.6%	60.2%
For phonemic awareness, teachers are explicit and systematic in their instruction.	34.5%	72.5%
Usually, teachers begin with auditory activities.	47.1%	75.3%

*Includes 618 teachers, specialists and coaches who had both pre- and follow-up surveys. The actual "n" for each item varies slightly due to missing cases.

It was relatively uncommon to observe lessons that focused on phonemic awareness – solely the *sound* components of words – without simultaneous work on phonics – the letter/sound connection. This was probably because site visitors observed classrooms toward the end of the academic year, by which time most programs had moved toward activities that combined phonemic awareness and phonics. Still, small portions of observed lessons did focus solely on student ability to manipulate sounds, as in this first grade classroom, where students practiced omitting onset sounds:

> Teacher: The word is "fall". Get rid of the /f/ and you have –
>
> *Students (together): All.*
>
> *Teacher: That's right. The word is "coat". Get rid of the /k/.*
>
> *Students (together): Oat.*
>
> *Teacher: Ok. Try this one. "Swing". Get rid of the /sw/.*
>
> *Some students: Ing.*
>
> *Other students: No! Get rid of /s/, the word is "wing".*
>
> *Teacher: You are right. I made a mistake. Good listening.*

Overall observed phonemic awareness activities tended to be fast paced and engaging, with students sitting on a rug around the teacher and responding chorally to questions or challenges from the teacher. In a few instances, site visitors who observed phonemic awareness lessons, especially longer ones directed to first grade students, questioned whether the students needed as much work with phonemic awareness as they were getting; the tasks seemed too easy and students appeared bored with them.

Phonics

Phonics instruction helps students learn the relationship between the phonemes (sounds) they hear in words and the graphemes (letters) they see written on the page. Students use their knowledge of those relationships in order to read as well as to write their own texts.

Even before implementation, the majority of surveyed staff believed that students needed explicit instruction in phonics. Table 10.6 shows very high disagreement with the statement that children do *not* require explicit instruction in phonics: among teachers, specialists and coaches, an average of 88.1 percent disagreed with this statement in the pre-test and 90.1 percent disagreed in the follow-up.

Table 10.6
Perceptions of the Need for Phonics Instruction (Survey Results)

In my experience, most children do not require explicit instruction in phonics.	Percent of Teachers*	
	Pre-Test	Follow-up Test
Strongly Agree/Agree	3.4%	3.6%
Strongly Disagree/Disagree	88.1%	90.1%

*Includes 618 teachers, specialists and coaches who had both pre- and follow-up surveys. The actual "n" for pre- and follow-up varies slightly due to missing cases.

Early phonics lessons work with students to explicitly link written letters and sounds. As students progress, lessons include greater complexity, including teaching students about exceptions and sight words as well as multiple spellings for the same sounds (such as using "s", "ss", "ce" or "sc" to represent the sound /s/). In fact, phonics lessons were frequently interwoven with spelling.

> *Teacher asks for the first sound in the words "cereal", "ceiling" and "celebrate".*
>
> *Teacher: So what do these words have in common?*
>
> *Student: They all have the /s/ sound at the beginning.*
>
> *Teacher: Yes. Does our language have a letter besides "s" that can make the /s/ sound?*
>
> *Student: The letter "c".*
>
> *Teacher: Does our language let the letter make more than one sound? (students nod). What sounds can "c" make?*
>
> *Students made /k/ and /s/ sounds.*

Like phonemic awareness lessons, work in phonics tended to be fast-paced and engaging. Regardless of the core reading program used in the school, site visitors saw students working with white boards or rearranging letter cards, which meant that most of the time, most students had an easy avenue for participation.

In some classrooms with teachers who were themselves non-native speakers of English, site visitors observed that teachers sometimes connected letters and sounds in patterns that were more typical of Spanish than of standard English. For example, a teacher lead first-grade students in an exercise that substituted one sound for another, and students needed to supply the letter to create the new word:

> *Teacher: Ok, take your word "mold" and change the /l/ to /s/. Then, change the /d/ to /t/.*
>
> *Students manipulate cardboard letters to make the words.*
>
> *Students: Most! Most!*
>
> *Girl: That was easy-peasy.*
>
> *Teacher: Ok, now change the /o/ to /e/. [pronounced like long "e" in "sweet"]. Some students are confused.*
>
> *Girl: Mist.*
>
> *Teacher: Yes. The letter "i" has two sounds, /i/ [like "eye"] and short /e/ [pronounced like long "e]).*

This teacher pronounced short "i" as though it were the long "e" of standard English (and consistent with Spanish pronunciation of the letter "i"). As these kinds of pronunciation issues surfaced in a number of observed classrooms, it might be helpful to work with coaches to help them make teachers more aware of accents, dialects and other variations, particularly in instances where the same letters make different sounds in Spanish and English.

Fluency

When applied to reading the term "fluency," encompasses not only the speed implied by the term, but also accuracy and phrasing. As students come to read more fluently, their attention is able to transfer from the task of decoding individual words to making meaning out of larger blocks of text.

In about one out of every five observation blocks, site visitors saw teachers working with students to improve the fluency of their oral reading. As fluency develops primarily through frequent practice at oral reading, many times observers saw teachers creating opportunities for students to read aloud, either to the teacher or to one another. Some teachers were skilled at

structuring fluency work to enable as many students as possible to practice at the same time – this effort to maximize engagement is discussed below.

While oral reading practice supports the building of fluency, by itself it does not ensure that students progress in accuracy or especially in phrasing. Since phrasing, or prosody, is important in conveying meaning clearly, some teachers focused on helping students use expression when they read. Some teachers reminded students that good readers did more than simply pronounce the words on the page, as this teacher told her class:

> Teacher: Hm, is that how we talk? (imitates a very flat reading)
>
> *Students: No! (laughing)*
>
> *Teacher: Ok, read it like we say it.*
>
> *Students read together with expression.*

Other teachers got the same message across by having students read with them or after them and modeling expressive reading. A site visitor saw one second-grade teacher manage this very successfully:

> *Teacher: Now we'll read chorally. I will read, and then you read. (She begins to read the text, with slightly exaggerated expression to illustrate that it is a question.) Do you ever wish you could climb a mountain?*
>
> *Students: (mimicking her expression and clearly registering that it is a question) Do you ever wish you could climb a mountain?*
>
> *They continued the pattern of teacher reading, students reading after her, continues for three minutes, then the teacher pauses to check for understanding.*

In a number of classrooms, grade-level fluency targets were posted on the board, often with charts tracking student fluency timings; it was evident that many students knew what their oral fluency goal was.

Vocabulary

One of the key findings of the National Reading Panel was that a knowledge of vocabulary and sufficient background information to comprehend were essential to successful reading. Explicit instruction in vocabulary—word meanings, an understanding of word parts, and an ability to ascertain meaning from context—can help students improve their reading ability.

Although the majority of teacher survey respondents (82.3%) believed it was "very true" that students were given direct instruction in both word meaning and word learning strategies,

vocabulary was the area most frequently mentioned by coaches during site visits as "in need of improvement". Findings from classroom observations provided some support for this impression, because although it was common to observe teachers and students working on vocabulary, the lessons were often less engaging than some of the others observed. Often, they covered words in only a superficial way. For example, in this first-grade classroom, the teacher wanted students to review words from the week's story:

> *Teacher holds up flash cards with vocabulary words: Afraid. Another word for scared is "afraid". Let's spell it.*
>
> *Students spell it together.*
>
> *Teacher. Spell it.*
>
> *Students spell it out together.*
>
> *Teacher: Next word is "bear", a big furry animal. Say it.*
>
> *Students: Bear.*
>
> *One student says "house", then laughs and says bear.*
>
> *Teacher: Spell it.*

In contrast, in one third-grade classroom an observer witnessed a strong vocabulary lesson in which the teacher used a range of techniques to encourage high levels of student engagement and a variety of approaches to practicing the vocabulary words of the day, "swiftly", "arranged", and "gazing":

> *Teacher reviews definitions, which are on the chalkboard and then puts a picture on the overhead.*
>
> *Student: Oh, I've seen that.*
>
> *Teacher: Yes, you have. We're thinking about the rich words we need to use in our writing. Look carefully. In this picture, what is moving swiftly?*
>
> *Student: Rabbit.*
>
> *Teacher (to all students): Thumbs up if you agree.*
>
> *Students put thumbs up.*
>
> *Teacher: What does it mean to move swiftly? Tell your partner. Give an example of something else besides a rabbit that moves swiftly.*

Students talk to partners at their tables.

Teacher draws a name on a stick from the can: Ellie, can you tell me what else moves swiftly?

Ellie: Cheetah.

Teacher: Yes, a cheetah moves swiftly. That means –

Students (together): fast!

Teacher (changing the picture on the overhead): Look at the picture. What is arranged in a neat pile?

Student: Fire.

Teacher: Is there fire there yet?

Student: No, firewood!

Teacher points to the word on the board: Arranged. Give your partner an example of something else that can be arranged.
Students consult each other. While students talk, teacher checks on one student, tells me it is his second day and he knows little English. He pays attention and listens to a pair.

Teacher draws stick from the can: Alex.

Alex: Books.

Teacher: Books can be arranged on the shelf…

The teacher continued this fast-paced lesson, using well-established classroom management routines to keep students on task even while they had a great deal of opportunity to talk.

Teacher: Name one animal that moves swiftly and one that doesn't. Think, think, think. (pause) Ok, talk, talk, talk.

Students turn to partner and give answers. Teacher circulates and listens to some of the answers.
Time to stop, she signals. One boy keeps on talking, and she motions with her finger to her lips. He stops.
…
Teacher: What is the difference between "glancing" and "gazing"? Think, think, think.. (pause) Talk, talk, talk.

Students talk. Most but not all of conversation is on topic.

Teacher: Robert?

Robert: [distinguishes clearly]

Teacher: Thumbs up if you agree.

Students put thumbs up.

Teacher: Which of these are arranged – sand on a beach or a tower of blocks? Think, think, think. Talk, talk, talk.

Students talk. One table is unsure.

Teacher (drawing a name on a stick): Mica?

Mica: The tower.

Teacher: Thumbs up if you agree. Osvaldo, do you agree?
Osvaldo nods.
Teacher: Now, what does arranged <u>not</u> mean? Think, think, think. (pause) Tell your partner.

Students, talking at the same time: Messy, not placed together, messed up.

Teacher: I am hearing some rich antonyms. Ok (draws name) Graciela?

Graciela: All over?

Teacher: Yes, all over the place, messy, or scattered. Write it down.

Ellie: Misplaced!

Teacher: Very good, you even used a word with "mis-", that's one of our prefixes. Write it down.

This lesson stood out not only for the strong classroom management skills of the teacher and high levels of student engagement, but also because in the course of about fifteen minutes, the students were led through exercises that had them not simply defining or memorizing words, but also generating their own examples, comparing the words to similar words, contrasting them with antonyms, restating the definition in their own words, and drawing a sketch to show the meaning. At the same time, vocabulary words from previous lessons lined the walls of the classroom, and students indicated by making "antennae" (wiggling fingers at the side of their heads) when they encountered those words, to indicate that their feelers were up and sensitized to rich vocabulary around them.

The less interesting lesson, more typical of what was observed in classrooms, probably adhered closely to the lesson plan that accompanied the core program, while the second example deviated in style, though not in content or spirit. The teacher in the second example later explained in an interview that she had been able to greatly enhance her vocabulary instruction using techniques presented in one of the professional development workshops she had attended that year. In the second year of implementation, as teachers become increasingly comfortable with the scope and sequence of the core program, there may be more room to encourage teachers to enrich the lessons in their teachers' manuals, perhaps especially in the area of vocabulary.

Comprehension Strategies

Research into practices employed by good readers to understand texts have identified certain strategies that can be explicitly taught to students to help them understand the texts they read. These include making connections between the text and students' lives and between new texts and previously read texts, or making predictions prior to reading, using graphic organizers to see the relationship between ideas, among others.

In many instances, however, teachers let very low-level comprehension questions, with an emphasis on recall, rather higher-order thinking, dominate the lessons observed. For example, a site visitor observed this comprehension lesson with a group of three third-grade students reading a story with the teacher:

Teacher: Why is he lonely?

Student: He lives in the desert.

Teacher: Who does he have as a friend?

Student: The donkey.

Teacher: According to the story, what do we look forward to?

Students do not answer, look confused.

Teacher: He is lonely. If he's lonely, what does he look forward to?

Student: Friends.

Teacher: What would he like to have happen?

Student: Have visitors.

Teacher: Does it sound like he gets many visitors?

Student: No.

Teacher: I want you to read the next page to yourself and be prepared to answer these two questions: what did he plant in his garden, and what is it he does after he plants his garden?

Similarly, in a first-grade class at another school, more than twenty minutes were spent on asking students, one at a time, to give brief answers recalling main events in the story:

Teacher: Where does Rosie live?

Students squirm on the floor.

Teacher: (waits) Does somebody know?

Student: She lived at an island.

Teacher: How do you call this island?
Several students: Jamaica!

Teacher: What was the main problem with the doll?

Student: Her arm was broken.

Teacher: What about Rosie's problems as the beginning of the story?

Student: She had a cold..

Teacher: What did the doctor say? (calls on student raising hand)

Student: Stay in bed.

Teacher: The doctor told her to stay in bed. So what was the solution to her problem?

Several students call out: Stay in bed.

Teacher points to a picture from the story and asks: What's happening?

Student: She's telling a story about her childhood.

Teacher (in Spanish): Estan de acuerdo?

Students: No! No!

Teacher: Think about it before you answer.

Students: It's a rag doll.

Teacher holds up a rag doll: This one we call "rag doll" because it is made with clothes.

In the latter example, both the teacher and the students were non-native speakers of English and both made frequent grammatical and pronunciation errors that did not interfere, however, with mutual understanding. The teacher's use of a visual prop, the rag doll she had brought into class, was a good example of a practice many teachers reported using to help their ELL students understand vocabulary that was new to them.

Comprehension lessons can involve a great deal more than asking students to answer simple questions about what happened or who the main characters in a story were. Teachers can also lead students to engage in higher-order thinking, for example by making connections between the text they read and other texts, or to their own lives or to the world around them. Teachers can also encourage students to predict, using picture cues or things they knew from reading part of a text, what might logically happen next, or to explain why, referring to passages in the text, they believe that characters behaved as they did.

These sorts of more challenging conversations about the meaning of text were observed in only a few classes. For example, in one third-grade class, the visitor observed a small group of students make predictions about a new story based on picture cues, read the story with the teacher – checking predictions as they went along – and then answer questions at the end. Rather than simply ask students to recall elements of the story, the teacher had them explain *how* the man in the story got from his position at the beginning of the story to the place he ended up at the conclusion, and she asked them to cite examples from the text to support their answer. This kind of questioning invited students to engage with text at a much deeper and more thoughtful level.

In a second-grade class at a different school, the site visitor observed as a teacher helped students make text-to-text connections, comparing the day's story to another story they had read previously:

Teacher: How could you compare bears from the two different stories we read, "The Bear Snores On" and "More About Master Snoozers"? (She makes a T chart on the blackboard.)

The students call out responses: furry, hibernating, snooze…

In interviews, coaches often told site visitors that teachers needed further opportunities to develop skills in teaching vocabulary and comprehension strategies. While the observations provided only short glimpses into classroom instruction, the relative infrequency of strong vocabulary and comprehension lessons observed supports the coaches' comments.

Systematic and Explicit Instruction

As the ADE noted in its original Reading First grant proposal, effective instruction is the single most important component of an effective reading program, and to be effective, instruction must be both systematic and explicit (Snow et al. 1998).

The structure of the core reading programs themselves, if the intended scope and sequence are followed, helps to provide the *systematic* presentation of sounds, letters, and concepts for beginning readers. The *explicit* piece, however, depends on the presentation of material from the teacher, as the same material from the identical reading program can be shared with students in different ways. Explicit instruction should include a substantial amount of teacher modeling. The "I do it, we do it, you do it" approach has the teacher demonstrating a task she wants to students to do, then doing it with students and then, as they grow more comfortable, backing off to allow them to do it on their own. In order to encourage higher level thinking about text, the teacher should also explicitly model the thinking process, in essence thinking aloud. To do this, teachers may say "I know that good readers predict, so I am going to think about what might happen next…" or "good readers make connections…".

The provision of systematic and explicit instruction was an area in which some schools felt they had made real gains over the past year. By the end of Year 1, there had been a noticeable shift in the percentage of schools where County Reading Specialists found at least "some evidence" that teachers consistently demonstrated appropriate, systematic, and explicit instruction in the five essential components. Specifically, 38.7 percent of schools were rated as "full implementation" of this item; 51.6 percent were rated as showing "some evidence" and only six schools (9.7%) – an improvement from 24 schools earlier in the year – showed "scant evidence."

These gains were also expressed during site visit interviews. For example, a teacher at one school felt the overall educational experience of students had improved because across the school there was a common approach to reading and it followed a logical and meaningful sequence:

> *Before the grant, every class did its own thing. Now there is emphasis and focus on objectives. It is all more organized. We make sure standards are taught in a sequential manner. (Teacher)*

Site visitors considered both the presence of explicit modeling and lesson clarity in the classrooms they observed. (Ratings are explained in the Methods section of this report.) In about half of the observed lessons, site visitors saw definite evidence of explicit modeling, and in 60 percent of lessons rated the lesson as "clearly presented" – the task, as well as the purpose of the task, was clear to the observer and appeared to be clear to the students as well see (Table 10.7).

Table 10.7
Observer Ratings of Lesson Clarity and Explicit Modeling

	Occasionally or Not at All	Yes, Definitely
The teacher models the work or thinking process.	49.3%	50.7%
Lesson is clearly presented.	39.7%	60.3%

In about half of the classroom observations, site visitors noted that the teacher provided good modeling of the reading and thinking processes they wished students to develop:

> *The teacher works with three kindergarten students, modeling sounding out a word. First she moves her own finger over the letters as she sounds out the single syllable words. Then she helps move the student's finger to guide it across the letters in a word, saying the word together with the student.*

In the other half of observed classrooms, modeling was weaker, more infrequent, or in a few cases, entirely absent. A number of teachers tended to tell students what to do, rather than to show them or do an example with them. It may be that by March or April there were enough established routines and familiar activities that teachers did not always need to model what they wanted students to do, but site visitors also observed looks of confusion from students who did not understand the task they were instructed to perform.

Occasionally, site visitors were surprised at the very explicit understanding students had of what they were doing. In one kindergarten class, a teacher reviewed punctuation with students and what a period or exclamation point meant for reading.

> Teacher: And why are we learning this?
>
> *Student: Because ADE says that kindergarten students need to read and write. This is a failing school because the big kids don't listen, but we won't be like them – we will learn.*

Other students were able to cite their fluency targets and knew that good readers built fluency so they could understand better what they read.

Using Time Effectively and Engaging Students

In addition to receiving high-quality, explicit instruction in the five essential components of reading, students need as many opportunities as possible to practice what they are learning. This requires both that teachers use classroom routines that reduce interruptions and wasted time, and that the protected time be structured in a way that as many students as possible are actively engaging with the material at the same time. According to County Reading Specialists' responses to the implementation checklist, schools improved on items related to student engagement during the year but still had room for growth. As shown in Table 10.8, the majority of schools showed "some" but not "full" implementation of actively engaging students 85

percent of the time and providing multiple opportunities for practice. According to CRSs, there was slightly higher evidence that teachers utilized activities that were respectful and purposeful for all students.

Table 10.8
Student Engagement (Implementation Checklist)

	May 2004 Percent (n)			Change from February*
	No/ Scant	Some	Full	
Teachers provide instruction in which all students are actively engaged at least 85% of the time.	4.8 (3)	74.6 (47)	20.6 (13)	+ 33 schools
Teachers provide multiple, varied practice opportunities in the five components of reading.	14.3 (9)	55.6 (35)	30.2 (19)	+ 11 schools
Teachers utilize instructional activities that are respectful and purposeful for all students.	4.8 (3)	57.1 (36)	38.1 (24)	+ 13 schools

*Number of schools that increased from "no" or "scant evidence" to "some evidence" or "fully implemented" from February to May. See text for discussion of limitations.

Site visitors also looked at several aspects of classroom management and student engagement, finding some excellent examples but also many opportunities for teachers to develop further in the effective use of time.

In the spring of 2004, reading was generally not interrupted by math, assemblies, or other activities. However, at some schools more minor interruptions by phone calls or requests from the office were still common. Sometimes the very location of a class made it hard to keep the students' undivided attention. One group of eight third-grade students met with the teacher in the hallway, where students were distracted by frequent passersby; in the space of a twenty-minute observation, nine students passed by to the go to the restroom, and another three students and two adults passed to enter the multi-purpose room.

A few teachers handled such interruptions especially well, bringing students quickly back to the task at hand, as in this second-grade class:

> *Teacher and 23 students are working on a vocabulary lesson, and students are highly engaged.*
>
> *The intercom interrupts; a student is to go to the office. At the same time, two students enter the class, late. Other students turn to them.*
>
> *Teacher: What happens if we have a distraction?*
>
> *Students: Just keep working!*
>
> *Teacher: Hold up the card with the word that rhymes with "wood".*

Students hold up their cards.

Probably a bigger concern than such interruptions, which were generally short, were the instances of inefficient transitions, when students lost potential learning time because it took so long to provide instructions or to get students from one activity to another. For example, in this second-grade classroom, about 12 minutes of a 20 minute observation were used in giving instructions and having students move from one group to another:

10:00 Teacher give instructions to whole group (10 minutes)

Teacher tells students instructions, then asks one group to read the instructions. Teacher waits for other students to listen to the group reading.

The teacher then explains the spelling/vocabulary center and gives instructions for the "Word Hunt" activity.

The teacher tells groups which center or activity to go to. Teacher makes sure students know where they are going.

10:10 Student move to small groups.

Students are completing transition to their centers/groups. The teacher waits for students to settle down.

Teacher (to small group of 6 students): Ok, open your book to the question list. Point to the first question and read it aloud.

[Lesson continues]

10:16 Bell rings

Teacher: Ok, that means the group time is up. You need to go to your new groups.

10:18 Students move to different small groups.

A new group of 5 students joins the teacher.

Teacher: Ok, go to the question list in your book. Read the first question together and then look for the picture that shows the answer…

While this amount of wasted time was not observed frequently, site visitors did note that even otherwise strong teachers had room to improve in the effective use of time. For example, in one kindergarten class where the site visitor saw wonderful modeling and highly engaging group work on vocabulary and phonics, there was also a loss of instructional time because some materials were not ready. In one class, the teacher spent six minutes organizing papers at her

desk for the next part of the lesson; while students practiced partner reading during that time, most finished before the teacher was ready and grew restless and fidgety.

On the other hand, site visitors also witnessed a number of positive examples of quick and efficient transitions, even with very young students. For example, one kindergarten teacher said she had spent time early in the year training her students to "rotate" quietly from one group to another, and the site observer noted the effectiveness:

> *8:55 Small groups*
> *Teacher works with a group of three students on decoding. The other 14 in the class are busy in their own groups.*
> *Teacher looks up: Ok class, get ready to rotate.*
>
> *8:56 Transition*
> *Teacher: Erase your white boards. Ready to rotate.*
> *Students stand at their places besides their desks. The teacher helps one group quickly put materials in their place.*
> *Teacher: Ok, rotate.*
>
> *8:58 Small groups*
> *Teacher models reading of one syllable words with a new group of four students. Other students are at their tables, already working.*

With established routines, this teacher was able to move very young students from one activity to another and have them working at a new task in under two minutes.

Another piece of student engagement is ensuring that work that students do when another group is with the teacher is meaningful and supports reading. Because the observations focused on the teachers and there instruction, there was little observation data collected to assess how well this was happening. It was, however, something that teachers from several schools mentioned as a real change from what they had done in the past:

> *I have always used centers, but they didn't necessarily go along with what we were studying. Now my learning centers are more connected to what we are teaching in reading.*
>
> *Last year, I taught reading for one hour… I had no centers and used chapter books with the whole class… This year I pair lower and higher students to help them partner and build vocabulary and they also work in centers while I am with a small group. I am able now to meet with all the levels.*

Whether in whole group, small group, or at centers, students needed ample opportunities to actively practice what they were learning. Site visitors noted that in over half of the classrooms observed (58.0%), teachers ensured that students were provided ample opportunities to practice the content of the lesson (Table 10.9).

Table 10.9
Observer Ratings of
Student Opportunities to Practice and Student Engagement

	Occasionally or Not at All	Yes, Definitely
Students have opportunities to practice the content of the lesson.	41.9%	58.0%
All students are engaged in the lesson.	51.4%	48.6%

Some activities, especially phonemic awareness and phonics lessons, easily gave all students opportunities to practice; when all students had white boards to write letters or words, or when kindergarten students had alphabet mats to use to point out the letters sung in a song, there was always something for every student to do. Perhaps for this reason, site visitors were more likely to see ample practice opportunities in kindergarten and first-grade classrooms, where there was more work on phonics and phonemic awareness.

When the lesson was focused more on talking about ideas, however, it was up to the teacher to provide students enough opportunities to practice. In one third-grade classroom, a teacher managed this by breaking her 21 students into five groups that worked together.

> *Teacher: We've been talking about whether or not these situations are realistic. Now I want you to discuss in your groups what realistic fiction is.*
>
> *Students talk in their groups.*
>
> *After a minute, the teacher counts down: Five, four, three, two, one.*
> *Students grow quiet again and look at the teacher.*
>
> *The teacher draws a popsicle stick from a can and calls on the student whose name was on the stick: What is realistic fiction?*
>
> *Student: It means something in the story that makes it true.*
>
> *Teacher: Something in the story seems to be true. What else did you talk about? (Draws another stick from the can and calls on another student).*

Later in the same lesson, the teacher gave each group a table to fill out with information from the story, and the students collaborated in completing the work. With this arrangement, and even with some off-task behavior during the latter part of the lesson, most of the time at least 22 students were actively involved and practicing the day's material.

Other teachers, by the very way they directed questions, limited students' opportunities to practice. For example, throughout an entire comprehension lesson one second-grade teacher repeatedly directed a question to the entire group of nineteen students, called on one student who raised his or her hand, and then moved on. Simply by asking students to share an answer with a partner would have greatly increased the number of students with an opportunity to provide a response to her questions.

One problem observed in some classrooms was the continued use of "round robin" reading. One site visitor observed a third-grade classroom spend the entire 20 minute observation engaged in "round robin" reading. In this arrangement, a group of students sit and read along silently as one student reads a paragraph or a page aloud. When that section is completed, the next student in the circle or row reads. This set-up makes it easy for the students who have already read to tune out, and this was certainly the case in that third-grade classroom:

> *Group of five students is reading aloud, round robin style, with the teacher (the other eleven are reading silently at their desks).*
>
> *The fifth student in the circle seemed very bored waiting for her turn. The teacher couldn't see her even though they were all seated at a horseshoe-shaped table; the teacher was turned to the first student, who was reading.*
>
> *When the student struggled with a word, the teacher told him the word without helping him sound it out. By the time this had happened a few times, students anticipated this and did not attempt a word they were unsure of.*

Although this was the only example site visitors reported of such an extended period of round robin reading, there were frequently other examples of many students sitting quietly – sometimes paying attention, sometimes not – while only one student at a time read aloud. In contrast, techniques such as "think – pair – share", which has all students share an idea with a partner before one or two share with the entire class, were rarely observed.

In rating student engagement, site observers adopted the same engagement target provided to principals at the 2003 Summer Academy – 85 percent of students should be engaged in the lesson. Just under half (48.6%) of observed lessons met this standard (see previous table).

Classroom Environment

The use of space and availability of materials are also tied to student engagement. CRSs rated schools on several implementation checklist items related to the physical environment of classrooms. At the end of Year 1, almost two-thirds of CRSs reported full implementation of adequate physical space for different instructional groupings and the presence of a wide variety of reading materials. However, only about half the schools have reached "full implementation" of the display of student work in classrooms or incorporating students' home cultures as shown in Table 10.10.

Table 10.10
Environment (Implementation Checklist)

	May 2004 Percent (n)		
	No/ Scant	Some	Full
A wide variety of engaging reading materials, both fiction and nonfiction, are available to students in reading classrooms.	7.9 (5)	28.6 (18)	63.5 (40)
K-3 reading classrooms are arranged to provide space for small group work, individual and partner reading, as well as whole group instruction.	3.2 (2)	34.9 (22)	61.9 (39)
Student work is prominently displayed in classrooms and on the school campus.	6.3 (4)	41.3 (26)	52.4 (33)
K-3 classrooms incorporate elements that support instruction and recognize students' home culture (books, posters, signs) in order to facilitate learning and make connections.	6.3 (4)	49.2 (31)	44.4 (28)
K-3 classrooms include word walls that teachers effectively incorporate into instruction.	19.0 (12)	60.3 (38)	20.6 (13)

Strategic and Intensive Interventions

For some students, especially those who start the year reading below grade level, systematic and explicit instruction from the core reading program may not be sufficient to help students make the necessary progress. For this reason, students who were identified on the DIBELS as reading below "benchmark" level were to receive targeted intervention – additional instruction targeting their very specific needs.

According to state project staff, the development of a comprehensive system to accurately identify students for intervention and then regularly deliver and monitor appropriate interventions was not the first priority for schools this Year 1. In the face of many other challenges – adopting a new core reading program, establishing a Reading Leadership team, developing collaborative structures and strong school leadership – interventions were left as a topic to receive more attention in Year 2 of implementation (2004-05). Not surprisingly, this meant that many schools struggled in year, confused about how to deliver what students needed. Some schools, especially those with large numbers of students reading below grade level, determined that "everyone needed interventions." By the end of the school year, state project staff knew this was an issue that would need to be addressed:

> *One thing schools do not always understand is that the small group instruction should not take the place of additional intervention for students who are substantially below benchmark. It is not the same as the 30 minutes a day of interventions.*
>
> *One of our biggest challenges with schools has been to define and to help them visualize and then implement a literacy block that includes the different pieces. Schools struggle with when, where, how should it occur, and especially what is intervention. They tell us they are hearing mixed messages when we say all students get instruction at grade level and in addition there is intensified instruction for students who are low. In some schools, there are so many students with difficulties, maybe up to 80 percent, that they think whole*

> *group is enough. Everyone needs help, so they give it to the whole group. They are not understanding that the intensified instruction needs to be very targeted to individual and specific needs identified through assessment.*

Because the ADE wanted schools to focus on learning and implementing their core programs during the first year, they did not offer professional development or resource lists of specific supplementary programs which could address the gaps in the core programs or provide materials for additional practice. Thus, it is not surprising that over half of the visited schools (13) reported that they did not have any supplementary materials. A few schools reported that teachers were, "doing their own thing" in terms of supplementing the core, and the remaining schools listed supplementary programs that already existed at the school or were obtained by the coach, principal, or district. Most schools knew that this would be a focus next year.

Despite the absence of a focus on interventions, CRSs rated the majority of schools as having at least "some" evidence of many aspects of intervention strategies and reported that at least nine schools had moved from "no/scant evidence" to at least "some" evidence by May 2004. For example, CRSs reported that 79.4 percent of schools showed at least "some" evidence of providing interventions in addition to the core reading program (12 schools more than in February).

However, the checklist ratings for items related to intervention were low compared to other categories in the checklist (Table 10.11).

Table 10.11
Intervention Strategies (Implementation Checklist)

	May 2004 Percent (n)			Change from February *
	No/ Scant	Some	Full	
Data are used to monitor the progress of students in intervention and adjustments are made accordingly.	7.9 (5)	38.1 (24)	54.0 (34)	+ 9 schools
Intervention is delivered by trained personnel, who demonstrate an understanding of SBRR and explicit, systematic instruction.	9.5 (6)	57.1 (36)	33.3 (21)	+ 13 schools
An appropriate and consistent allotment of time has been designated for intervention instruction weekly.	14.3 (9)	39.7 (25)	46.0 (29)	+ 15 schools
Intervention instruction (Tier Two and Tier Three) is provided in addition to the 90-minute minimum instruction in the core reading program.	20.6 (13)	41.3 (26)	38.1 (24)	+ 12 schools
Purchased intervention materials/programs are used appropriately.	41.3 (26)	31.7 (20)	27.0 (17)	+ 9 schools

*Number of schools that increased from "no" or "scant" to "some" or "fully implemented" from February to May. See text for discussion of limitations.

During site visits, school staff members and assessment coordinators were queried about their intervention systems in terms of monitoring, successes, and challenges. Responses from these interviews highlighted the fact that interventions were not a focus of training during the year as challenges were identified in regard to developing intervention systems, selecting materials, and providing adequate staffing.

Intervention systems were clearly in different stages of development at different schools. A few schools acknowledged they had no system in place and/or needed to develop an intervention system, and like the ADE, intended to focus on this in the coming year.

> *Teachers are not into interventions yet and there is no systematic 30 minute intervention implemented at the school. (Coach)*

Other schools reported they were in the development stages. Some schools were using the core curriculum's intervention program:

> *The only interventions used are part of the core program. While I guide teachers on what they should focus on, in the absence of a true intervention program, teachers do what they want. I do not monitor the interventions. (Coach)*

In other schools, some group (e.g., RLT or a curriculum committee) provided input into the delivery of interventions, or interventions were planned by a team comprised of some combination of teachers, the reading coach, the assessment coordinator and/or Title I/ELL teachers.

> *The Reading Leadership Team had a say in how interventions were monitored and which students received support; they decided that at least five minutes would be spent on phonemic awareness, at least five minutes on phonics, and ten minutes on guided reading. (Site visitor)*

> *The assessment coordinator and reading coach monitor the test scores where they look for stagnating scores. When that arises they conference with the teachers on potential strategies. According to the coach, if [scores are not improving], this is red flag to us to take action. (Site visitor)*

Finally, a few schools appeared to have made progress in developing intervention systems, such as one which was described by an evaluator as follows:

> *The Reading Lab has been constructed by the school as a separate entity. It has 4 stations manned by aides working with small groups of students. It is monitored either by the Coach or the Assessment Coordinator. As a team, the Assessment Coordinator and the Coach take a look at the progress being made by the student after two weeks at the Reading Lab. (Site visitor)*

Not only were schools struggling with the development of a system for serving at-risk readers, they were also struggling with the identification of materials they could and should use. Some schools had not yet identified materials that might be used to provide interventions. For example, a group of teachers on one RLT requested "intervention programs ready to go" so they would not have to spend time "digging" for materials. At another school a package had a least been identified, but staff members were not sure if they could use it. A teacher on that RLT asked:

> *Can I use Zoo Phonics as an intervention? Or can I only use the core? (Teacher)*

Some schools, as noted earlier, were using their core program's intervention materials. In some cases these materials were meeting their needs, they were coordinated with classroom lessons, and/or they gave teachers additional methods for teaching. At other schools, however, the core program's intervention materials were clearly not meeting the needs of the schools' struggling readers.

> *There is no systematic 30 minute intervention being implemented at this school. According to the RLT, core program intervention materials are not strong enough and so staff plans to develop an intervention system next year. (Site visitor)*

> *The intervention strand of the core program does not meet the specific needs of the school's struggling students. (Site visitor)*

Some schools had successfully identified pre-existing intervention packages to use in addition to or instead of core interventions (e.g., CLIP, HOST, Great Leap, Spalding) that "really help to work with struggling readers," "give an assessment and prescription," or fill a specific need such as phonemic awareness. Others still wanted more information on what to do with struggling

readers (i.e., those not responding to initial interventions, ESL/ELL students, and special education students), more lower-level materials, and more materials in general. One evaluator summarized the sentiments of the coach and assessment coordinator as follows:

> *They are happy with the general structure of the system, but want to know more about specifically how to work with students who are not getting it. (Site visitor)*

Staffing of interventions was the most frequently cited challenge in providing interventions—schools did not feel they had enough staff to provide and monitor the necessary interventions. Schools expressed a need for additional staff members because students were served at a variety of times (e.g., before- and after-school, summer school), because effective interventions required differentiated/targeted support, and because the number of students who were targeted for interventions continued to grow.

> *There are still not enough staff to meet the needs of all the struggling students. The teaching staff is facing severe burnout fatigue with just not enough personnel to meet the current needs. (Site visitor)*

In almost all visited schools (95%), certified teachers provided interventions. In addition, paraprofessionals provided some interventions in two-thirds of the visited schools; coaches shared the responsibility in 60 percent of the schools; volunteer/paid tutors had this responsibility in one-third of the schools; and other staff members were reported as providing interventions in about one-quarter of the schools. Several staff members also noted that staffing was a problem because aides were not allowed to work with students and because students were not allowed to be pulled from the classroom.

Ironically, as intervention programs raised concerns for some teachers, they were identified as solving problems for others: some teachers reported they no longer felt solely responsible for the struggling readers in their school or that the responsibility of developing and/or implementing an intervention was being placed solely on their shoulders. They also helped to ensure consistency in the planning and implementation of interventions.

> *It's nice to have extra people to work with these kids, I feel than now I have a team helping me with these kids and before it was just me. (Coach)*

Miscellaneous comments from staff members indicated that staff development might be useful in the following areas: increasing buy-in to an intervention program, interpreting and using DIBELS data, effectively differentiating intervention lessons, defining interventions and delineating what materials could be used to provide the most appropriate interventions.

Meeting the Needs of English Language Learners

By law, Arizona classrooms are English-only and the ADE expected schools to follow the English immersion guidelines that exist in policy and statutes. All students are expected to read and write in English, and have to be taught and assessed in English. Together with native English speakers, all ELL students were to receive at least 90 minutes of uninterrupted reading

instruction in English every day. Outside the 90-minute block, ELL students should received additional instruction in oral language development to aid in language acquisition.

All 23 of the visited Arizona Reading First schools reported that they served ELLs, though the proportion of ELL students in the school ranged from one to 100 percent. While ELL students made up only a small percentage of students in a few schools, most schools reported a population of at least 20 percent and almost half of the schools reported ELL students made up the majority of their student population. These high percentages across the schools make clear the importance of a reading program that is effective with ELL students.

Both observations and interviews confirmed that, indeed, ELL students received regular reading instruction, in English and with a core reading program, during the 90-minute block. While interviewed staff indicated their core programs were generally useful in their standard format, many schools also used the ELL components that accompanied their core reading program materials. Respondents expressed differing views on how well the core program met the needs of ELL students, and many noted the need for additional focus in certain areas for ELL students, especially vocabulary.

> *For those with limited English, the core is not so good because the level of the materials are too high for them. (Teacher)*
>
> *I love the materials in the core, however, it remains impossible to meet all of the needs of all the students with any set curriculum (Teacher)*
>
> *We have an ELL program as part of the core program. They have an extra support handbook that has a vocabulary component that is good, and it's used during interventions if needed. Modifications that are used come from the curriculum. (Teacher)*

Before grant implementation, survey respondents were only somewhat confident that they could describe and model effective strategies to work with ELLs; respondents were evenly split between those who chose "very true" (46.2%) and those who chose "somewhat true" (40.2%) to this item. By the spring, the percentage of teachers, specialists, and coaches choosing "very true" had increased to 62.4 percent and the percent of respondents who felt it was "not true" decreased from 13.7 percent to 2.3 percent. However, over one-third of respondents still felt it was "somewhat true," indicated a continuing need for training.

Table 10.12
Supports for English Language Learners (Survey Responses)

Statement: I could describe and model effective strategies to work with English Language Learners.	Percent responding*	
	Pre-test	Follow-up test
Very true	46.2%	62.4%
Somewhat true	40.2%	35.3%
Not true	13.7%	2.3%

*Includes 618 teachers, specialists and coaches who had both pre- and follow-up surveys. The actual "n" for pre-test and follow-up varies slightly due to missing cases.

During interviews with teachers whose classrooms were observed, two-thirds of those who worked with ELL students were able to explain specific modifications they made to better meet student needs. An additional twenty percent described vague modifications, and the remaining ten percent of teachers did not explain any modifications or said that modifications were not needed.

The modifications that teachers most frequently mentioned were using visuals (e.g., pictures or charts) and using physical experiences, such as hands-on activities, movement, or acting out.

> *Constantly I make modifications…I use acting a lot, I pantomime a lot, through the use of actions and visuals I try to help them. (Teacher)*

> *I also try to bring pictures, role play, and use acting out to help with prior knowledge. Other kids have an opportunity to be little teachers by translating to their fellow students. (Teacher)*

Other modifications included pairing students, using music or chants, and having another student translate for a struggling ELL student. Some teachers reported that they could incorporate these modifications into whole-group instruction without singling out ELL students. Other teachers reported that paraprofessionals worked with some ELL students when it was time for small group work.

Despite modifications like those described above, many interviewed staff felt that working with ELL students within the structure of the Reading First program was a challenge. For example, one interviewee said, "There is a lack of a strong ELL component in Reading First." Another interviewee said, "The core is not working for ELL students." Staff wanted more resources, materials, and strategies for making progress with the ELL population.

> *We always need some ELL help…it is very quiet at the state and national level about truly how to teach them. Evidently there is little research, so they say "use your core," and we do the best we can, but we need someone to do some heavy duty research in the area. (District representative)*

Outside of instruction during the 90-minute block, site visitors did not specifically inquire about the details of additional supports for ELL students. However, many schools offered up their own lists of additional supports they used to meet student needs, including the following:

- ESL professional development provided for teachers by the school or district
- ELL-endorsed / ESL–certified staff in classrooms
- Bilingual teachers and/or paraprofessionals
- Additional instructional support during the school day or before/after school
- Programs such as Sheltered English Immersion
- English programs for parents
- Additional classes for ELL students such as language development provided when school was not in session

Only one school reported that they pulled out students for ELL classes. Another school reported that, because of a civil rights lawsuit charging that ELL students were missing instruction from teachers because they were pulled aside with less-qualified aides, the district does not permit ELL pull-outs and the school has no ELL specialists. Instead, all teachers were expected to know how to work effectively with English Language Learners.

REFERENCES

Foorman, B.R. and C. Schatschneider. (2003). Measurement of teaching practices during reading/language arts instruction and iIts relationship to student achievement." In Vaughn, S. and K.L. Briggs, *Reading in the classroom: systems for the observation of teaching and learning.* Baltimore: Paul H. Brookes.

Fouts, J.T. (2003) *A decade of reform: A summary of research findings on classroom, school, and district effectiveness in Washington state. Research Report #3.* Lynnwood, Washington: Washington School Research Center, Seattle Pacific University.

McCarthy, M.S. & Celio, M.B. (2001). *Washington state elementary schools on the slow track under standards-based reform.* Seattle: Center on Reinventing Public Education, Daniel J. Evans School of Public Affairs, University of Washington.

Snow, C E., Burns, M.S & Griffin, P. (1998). *Preventing reading difficulties in young children.* Washington DC: National Academy Press.

Taylor, B. M., Person, P.D., Clark, K. and Walpole, S. (2000). Effective schools and accomplished teachers: lessons about primary-grade reading instruction in low-income schools, *The Elementary School Journal.* Vol. 101, Number 2.

Washington School Research Center. (2002). *Bridging the opportunity gap: How Washington's elementary schools are meeting achievement standards. Research report #2.* Lynnwood, Washington: Washington School Research Center, Seattle Pacific University.